

Suomen hallinnollisten alueiden muutosten kuvaaminen ja hyödyntäminen tiedonhaussa

Jari Väättäinen

Sähkötekniikan korkeakoulu

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi
diplomi-insinöörin tutkintoa varten

Espoossa 25.1.2016.

Työn valvoja:

Prof. Eero Hyvönen

Työn ohjaaja:

Dos. Tomi Kauppinen

Tekijä: Jari Väättäinen		
Työn nimi: Suomen hallinnollisten alueiden muutosten kuvaaminen ja hyödyntäminen tiedonhaussa		
Päivämäärä: 25.1.2016	Kieli: Suomi	Sivumäärä: 6 + 70
Tietotekniikan laitos		
Professori: Mediatekniikka		Koodi: CS.thes
Valvoja: Prof. Eero Hyvönen		
Ohjaaja: Dos. Tomi Kauppinen		
<p>Paikkatiedon merkitys tietoaaineistojen löydettävyydelle on suuri. Lähes kaikkeen toimintaan liittyy paikkatietoa, jota tallennetaan tietojärjestelmiin yleisesti paikannimien avulla. Paikannimistön käyttöön liittyy kuitenkin ongelmia, koska se ei useinkaan yksilöi paikkoja riittävästi ja varsinkin hallinnollisten paikkojen nimistö ja rajat muuttuvat melko usein ajan kuluessa. Tässä työssä esitellään Suomen hallinnollisissa alueissa havaitut 14 erilaista muutostyyppiä ja selvitetään, miten ne vaikuttavat paikannimistön perusteella kuvaillun tiedon löydettävyyteen. Työ pohjautuu Suomen hallinnollisissa alueissa tapahtuneiden muutosten mittavaan kartoitukseen, jossa löydettiin yli 1100 alue- ja nimimuutosta aikavälillä 1865–2015.</p> <p>Työssä kuvataan tesarus, GIS ja ontologiamenetelmiä, joiden avulla ongelmia voidaan ratkaista. Tarkemmin esitellään tietokannan muutostaulua hyödyntävä menetelmä www-palvelussa, logiikkaan pohjautuva menetelmä valokuvatietokannassa sekä Suomen ajallinen paikkaontologia (SAPO). Työssä on laskettu, miten Suomen kuntien avulla kuvaillun aineiston löydettävyys muuttuu ajan suhteen koko kuntien olemassaolon aikana vuodesta 1865 alkaen. Aineiston saantia ilman muutostenhallintaa verrataan saantiin muutostenhallintaa käytettäessä. Aineiston löydettävyys paranee merkittävästi jo pelkästään hallinnollisten alueiden nimenmuutokset ja yhdistymiset huomioimalla, joka on helppoa sekä ontologiaa että reaaliotietokantaa käytettäessä.</p> <p>Työssä esitellään myös paikkatiedon hallintaa koskevan useissa suurissa suomalaisissa organisaatioissa tehdyn kyselyn tulokset. Tärkeimpinä tuloksina kyselyssä saatiin, että muutostenhallinnalle on kaikissa kyselyyn vastanneissa organisaatioissa havaittu olevan tarvetta ja että tietojärjestelmien käyttäjille tulisi tarjota mahdollisuus hakea aineistoa sekä alkuperäisen että uusimman kuntajaon mukaisilla paikannimillä.</p>		
Avainsanat: aika, ajallinen paikkatieto, hallinnolliset alueet, muutos, paikkatiedot		

Author: Jari Väätäinen

Title: Modeling of the changes in Finnish administrative areas and using the models in information retrieval

Date: 25.1.2016

Language: Suomi

Number of pages: 6 + 70

Department of Computer Science

Professorship: Mediatechnology

Code: CS.thes

Supervisor: Prof. Eero Hyvönen

Advisor: Dos. Tomi Kauppinen

Spatial data has a big influence in information retrieval. Almost all types of information contain some kind of spatial data which is saved in data systems usually using place names. However, using place names has some problems. One problem is how to disambiguate different places with the same names. Another is that the names and borders of administrative areas are changing over time. In this thesis fourteen different types of changes in Finnish administrative areas are represented. It is also explained how changes affect the findability of those places in information systems. This work is based on a big survey of the changes in Finnish administrative areas where over 1100 areal and name changes were found.

In this thesis gazetteer, thesaurus, GIS and ontology methods are explained and used to resolve the mentioned problems. Three different kinds of methods are more specifically defined. They are a www-service where a change table in a database is used to handle changes in administrative areas, a method in a photo database that is based on logics and the Finnish Spatio-temporal Ontology (SAPO). It is calculated how much the recall of material annotated with Finnish municipalities is changing through time when searched with current municipalities. The recall values without using any method are compared with the values calculated when using each method. As a result was found that the recall has a remarkable raise even using only the name changes and data about the mergers of the administrative areas. This method is easy to implement using an ontology or as in a relational database with a change table.

This thesis also presents the results of interviews about spatial data management in six organizations in Finland. The most important results of the interviews were that the organizations see a big need for a method for managing spatio-temporal data and that there should be a possibility to search both with the original and with the present place names in the data systems.

Keywords: administrative areas, change, geographic information retrieval, spatio-temporal data, time

Alkusanat

Kehittäessäni Geologian tutkimuskeskuksen valokuvatietokantaa huomasin monien kuvauspaikkojen sijaitsevan kunnissa, joita ei enää ole olemassa. Päädyin miettimään ratkaisua siihen, miten kuviin tallennettu alkuperäinen paikkatieto voitaisiin säilyttää ja kuitenkin hakea kuvia myös paremmin tiedossa olevilla nykyisillä kunnilla. Ratkaisu löytyi ja päädyin keräämään muuttuneita kuntatietoja parin vuoden päästä GTK:n lisäksi Semanttisen laskennan tutkimusryhmälle, jossa kehitettiin Suomen ajallista paikkaontologiaa.

Haluan kiittää professori Eero Hyvöstä kipinän antamisesta opintojen aloittamiseksi Aalto-yliopistossa ja tuesta sekä kannustuksesta opintojen loppuun saattamiseksi – viimeisenä vaiheena tämän työn valvojana.

Kiitän myös diplomityön ohjaajana toiminutta dosentti Tomi Kauppista ajallisen paikkatiedon hallintaan liittyvistä monista kiinnostavista keskusteluista ja tämän työn toteuttamiseen liittyvistä neuvoista.

Lopuksi haluan vielä kiittää tämän työn yhteydessä tekemääni kyselyyn vastanneita henkilöitä ja organisaatioita.

Espoo 25.1.2016

Jari Väättäinen

Sisällysluettelo

1	Johdanto.....	1
2	Taustaa.....	6
2.1	Maantieteellinen tiedonhaku	6
2.2	Nimistöluettelot.....	7
2.3	Ontologiat.....	10
2.4	GIS-menetelmät	11
3	Muuttuva paikkatieto Suomessa.....	14
3.1	Suomen hallinnollisissa alueissa tapahtuneet muutokset.....	14
3.2	Muutosten vaikutus paikannimillä tapahtuvien hakujen tuloksiin.....	22
3.3	Muita syitä muutokseen alueiden pinta-aloissa.....	25
4	Muutostietojen mallintaminen ontologiassa.....	28
4.1	Muutoksia kuvaava ontologiasarja.....	28
4.2	Spatiaaliset muutossillat.....	29
4.3	Sijainnin todennäköisyyden laskeva päättelykone.....	30
4.4	Hallinnollisten alueiden rajapolygonit	31
4.5	SAPOn ylläpito ja julkaiseminen	32
5	Muutostietojen mallintaminen valokuvatietokannassa	36
5.1	Geologian tutkimuskeskuksen valokuvatietokanta	36
5.2	Muutosketjujen kuvaaminen dynaamisilla kategorioilla	36
5.3	Käytännön kokemuksia kategoriamenetelmän toimivuudesta.....	43
6	Muutostietojen mallintaminen verkkopalvelussa	44
6.1	Hakku-hakupalvelu	44
6.2	Valokuvat Hakku-palvelussa	44
6.3	Kuntamuutokset Hakku-palvelussa.....	46
7	Tulokset.....	49
7.1	Tarkkuus- ja saanti ilman muutostenhallintaa.....	49
7.2	Tarkkuus- ja saanti kategoriamenetelmää käytettäessä.....	50
7.3	Tarkkuus- ja saanti tietokannan muutostaulua käytettäessä.....	51
7.4	Tarkkuus ja saanti SAPO-ontologiaa käytettäessä.....	53
7.5	Ajallisen paikkatiedon hallintaan liittyvä kysely	54
8	Pohdinta.....	58
8.1	Tutkimuksen tuloksista	58
8.2	Katsaus tutkimuskysymyksiin.....	59
8.3	Tulevaisuudennäkymiä	60
9	Yhteenveto.....	62
	Lähdeluettelo	63

Lyhenteet

ADL	Alexandria Digital Library
Allärs	Allmän tesauros på svenska
AUO	Administrative Unit Ontology
BFO	Basic Formal Ontology
GCS	Gazetteer Content Standard
ETRS	European Terrestrial Reference System
GBHGIS	The Great Britain Historical Geographical Information System
GIR	Geographic Information Retrieval
GIS	Geographic Information System
GML	Geography Markup Language
GPS	Global Positioning System
GTK	Geologian tutkimuskeskus
HIPLA	Historiallisten paikkojen kansallinen nimistöpalvelu
IIM	Information Interchange Model
IPTC	International Press Telecommunications Council
IR	Information Retrieval
Kotus	Kotimaisten kielten keskus
MWG	Metadata Working Group
ONKI	Kansallinen ontologiakirjastopalvelu
OWL	Web Ontology Language
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SAPO	Suomen ajallinen paikkaontologia
SeCo	Semantic Computing Research Group
SKOS	Simple Knowledge Organization System
TGN	Getty Thesaurus of Geographic Names
TM35	Transverse Mercator, kaista 35
TURTLE	Terse RDF Triple Language
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
WGS84	World Geodetic System
XSLT	Extensible Stylesheet Language Transformations
XMP	Extensible Metadata Platform
YSA	Yleinen suomalainen asiasanasto
YSO	Yleinen suomalainen ontologia

1 Johdanto

Paikka on käsitteenä monitasoinen. Ruokahuoneen pöytä, koti, naapurusto, kaupunki ja valtio ovat kaikki paikkoja. Tuanin [47] mukaan paikka voi toisaalta tarkoittaa puhtaasti tiettyä spatiaalisesti määriteltä kohdetta ja sen ominaisuuksia tai toisaalta ihmisen johonkin tilaan liittyviä kokemuksia ja tuntemuksia. Paikkoihin liittyy ihmisillä siis paljon muitakin ominaisuuksia kuin sijainti- ja pinta-alatietoa – niihin voi liittyä myös vahvoja tunteita.

Paikkatieto (engl. *geographic information, spatial data*) tarkoittaa ”kaikkea tietoa kohteista, joiden paikka Maan suhteen tunnetaan” [36]. Jotta paikkatietoa voidaan tallentaa, tarvitaan metatietomääritteitä, jotka kytkevät tiedon ja paikan toisiinsa. Tällaisina määritteinä on yleisesti käytetty paikannimiä ja koordinaatteja. Tiedon hakeminen tietojärjestelmistä vastaavasti tapahtuu syöttämällä hakukenttään paikannimiä tai rajaamalla haluttu alue karttakäyttöliittymästä.

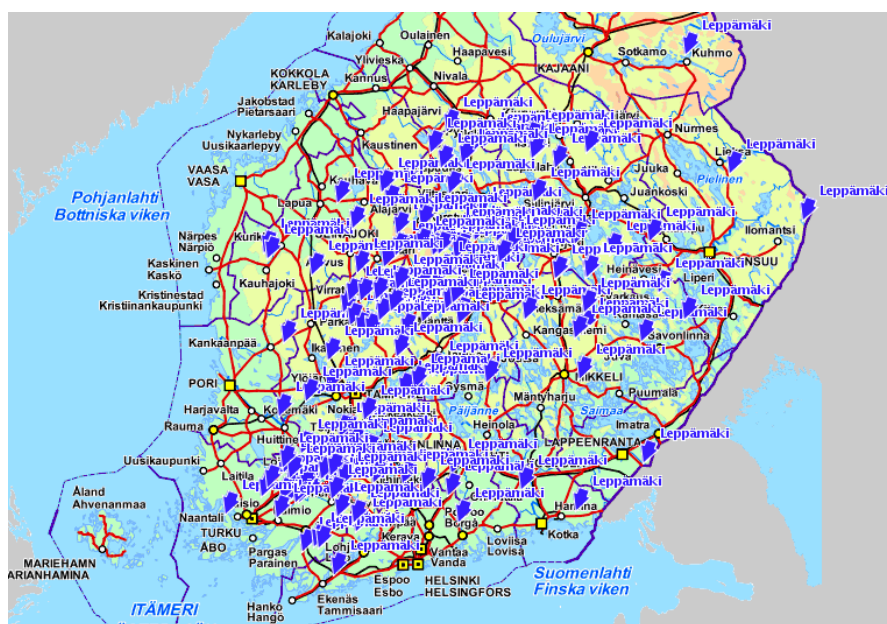
Paikannimistö on ollut perinteisesti ylivoimaisesti eniten käytetty tapa määrittää paikkoja [9, 16, 50]. Lehdissä, kirjallisuudessa, erilaisissa kokoelmätietokannoissa, tilastossa ja arkistoissa käytetään lähes yksinomaan paikannimiä paikkatietona. GPS-paikantimien ja koordinaattitietoja tukevien sovellusten yleistymisen muun muassa matkapuhelimissa on vauhdilla lisännyt myös koordinaattitietojen tallennusta esimerkiksi valokuvien ottopaikoista. Toki koordinaattien lisääminen on mahdollista jälkikäteen myös vanhempaan aineistoon karttapohjalta poimimalla, mutta se on varsin työlästä. Siksi sekin tehdään usein koneellisesti georeferoimalla paikannimien pohjalta. Onkin olemassa tarve menetelmille, joiden avulla tietojärjestelmiin tehtäviin paikannimiä käytäviin hakuihin saadaan tuloksena mahdollisimman relevantteja osumia [52].

Paikkatietojen tallentamista varten tietokannoissa on yksinkertaisimmillaan yksi kenttä, johon täytetään vapaamuotoisesti kaikki tiedossa olevat kohteeseen liittyvät paikkakäsitteet. Yleisempää kuitenkin on, että eri paikkatyypeille on omat tallennuskentät, esimerkiksi: maa, suuralue, kunta ja tarkempi paikka. Tarkempi paikka voi olla vaikka kylä, kaupunginosa, rakennus tai luonnon paikka. Esimerkiksi museotietokannoissa paikoille määritellään myös erilaisia rooleja: valmistuspaikka, käyttöpaikka, löytöpaikka, sijaintipaikka ja hankintapaikka [19, 46].

Paikannimen perusteella tietoa haettaessa ongelmaksi voi muodostua se, ettei paikannimi riitä kohdistamaan hakuja riittävän tarkasti oikeaan kohteeseen [52]. Yksi syy tähän on se, että paikannimiä vastaavia kohteita on esimerkiksi tuotemerkkien, organisaatioiden, henkilöiden ja yritysten nimissä. Esimerkiksi Nice tarkoittaa englannin kielessä eteläranskalaisen kaupungin (suomeksi Nizza) lisäksi mm. ystävällistä ja kaunista. Nice on myös suuri kansainvälinen yrityskonsultti. Suomessa taas esimerkiksi haku paikannimellä Nokia tuottaa tulokseksi paljon viittauksia myös Nokia-yritykseen.

Toinen vielä tavallisempi ongelma on samannimisten paikkojen runsas esiintyminen [17, 20, 50]. Esimerkiksi Paris-nimisiä paikkoja maailmassa on kymmeniä [50]. Suomessakin on runsaasti keskenään samannimisiä paikkoja [21]. Paikkatyyppien tarkka erottelu auttaa jonkin verran löytämään oikean kohteen samannimisten mutta erityyppisten paikkojen joukosta. Vaikka paikkatyyppi otetaan huomioon, on edelleen vaikeuksia erottaa kaikkia paikkoja pelkän nimen perusteella. Esimerkiksi Leppämäki-nimisiä paikkoja on Kansalaisen karttapaikan¹ mukaan Suomessa 134 kpl. Niistä 105 kpl on taloja, 20 kpl kohoumia, 6 kpl kyliä ja 3 muita maastokohteita. Kuvassa 1 on niiden alueellinen jakaantuminen kartalla.

¹ <http://kansalaisen.karttapaikka.fi/>



Kuva 1. Leppämäki-nimisten paikkojen sijainteja Suomen kartalla. Kartta: Maanmittauslaitos, Kansalaisen karttapaikka

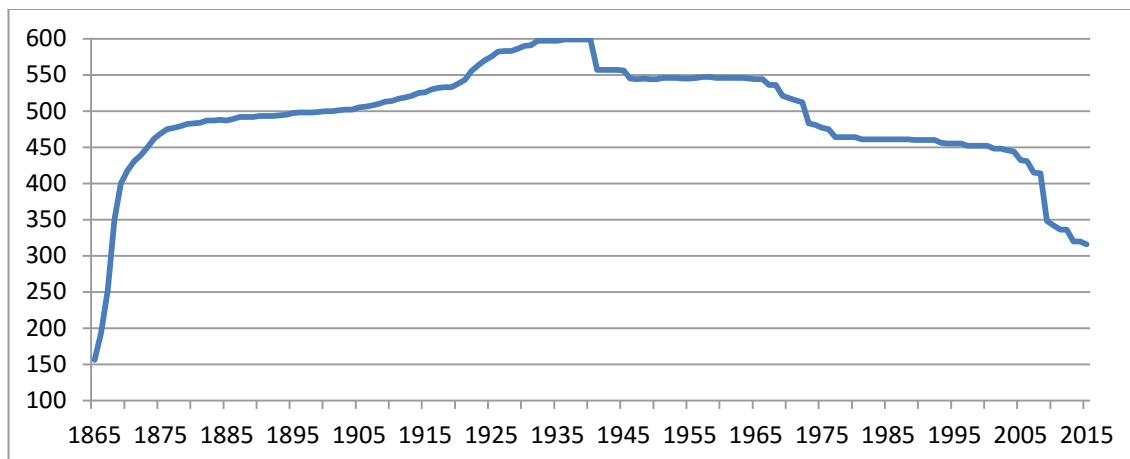
Myös hallinnollisten alueiden nimissä on ollut päällekkäisyyksiä. Ongelmien välttämiseksi käytetään paikannimen yhteydessä joskus tietoa laajemmasta hallinnollisesta alueesta. Esimerkiksi Suomessa on ollut kaksi Koski-nimistä kuntaa: Koski Hl. ja Koski Tl., ensimmäinen Hämeen läänissä ja toinen Turun läänissä [55]. Pyhäjärvi-nimisiä kuntia on Suomessa ollut yhtä aikaa jopa kolme. Myös ne erotettiin toisistaan nimeen lisätyillä läänipäätteillä. Vielä hankalampi tilanne oli, kun samassa Turun ja Porin läänissä oli 1800-luvun lopussa kaksi Eura-nimistä kuntaa. Nämä erotettiin kutsumalla toista “Euran kappeliksi” tai “läänin eteläisemmäksi Euraksi”. Yhtenä teemana tässä työssä käsitellään keinoja yksilöidä samannimisiä paikkoja.

Vanhoja kohteita tietokantaan tallennettaessa huomataan, että paikannimistö on muuttunut ajan kuluessa [20, 28, 40, 42]. Jopa valtioiden nimet ovat historian kuluessa muuttuneet: entinen Burma on nykyään Myanmar, Ceylon Sri Lanka ja Siam Thaimaa. Lisäksi on tapahtunut erilaisia aluemuutoksia. Neuvostoliitto, Tšekkoslovakia ja Jugoslavia ovat jakaantuneet. Saksa on ensin jakaantunut ja myöhemmin taas yhdistynyt. Historiallisia karttoja tutkimalla huomataan, että paikannimissä ja hallinnollisissa alueissa on kautta aikojen tapahtunut erilaisia muutoksia.

Myös Suomen alueella on tapahtunut sekä valtioiden välisiä että sisäisiä hallinnollisten alueiden muutoksia. Suomen hallinnollinen aluejako perustuu kuntajakoon [28]. Lukuisat tilastotiedot on koottu jaettuna kunnittain. Luonnontieteellisiin havaintoihin, valokuviin, museoesineisiin ja tapahtumien sijainniksi on merkitty vähintään kuntatieto, jos se on ollut suinkin tiedossa. Kunnan nimi onkin Suomessa yleisesti käytetty paikannimäire erilaisten aineistojen metatiedoissa. Esimerkiksi Geologian tutkimuskeskuksen valokuvatietokannassa on kuntatieto yli 99 prosentissa kaikista kuvista, joista on jotain paikkatietoa.

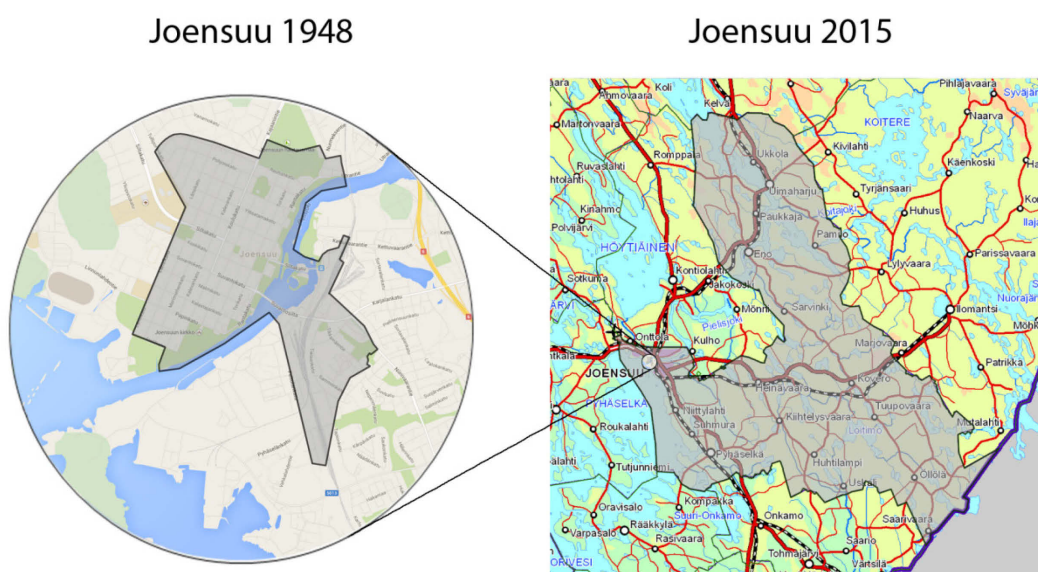
Kuntien lukumäärä Suomessa on vaihdellut runsaasti. Kuvasta 2 nähdään, että lukumäärä kasvoi 1930-luvun loppuun saakka, kun uusia kuntia perustettiin ja vanhoja kuntia jaettiin pienemmiksi. Sen jälkeen kuntien lukumäärä on vaihteellain laskenut. Laskussa on tapahtunut kolme nopeampaa jaksoa. Ensimmäinen sijoittuu talvisodan jälkeisiin alueluovutuksiin, jossa Neuvostoliitolle luovutettiin 45 kuntaa kokonaan ja 21

kuntaa osittain. Toinen lasku tapahtui 1970-luvun alussa ja kolmas kaikkein rajuin pudotus 2000-luvun ensimmäisen vuosikymmenen loppupuolella. Tuolloin esimerkiksi vuonna 2009 kuntien määrä väheni yhdessä vuodessa 67:llä. Kunniksi näissä luvuissa, samoin kuin tässä työssä yleensäkin, on laskettu myös kaupungit.



Kuva 2. Kuntien vuosittaiset lukumäärät Suomessa. Lukumäärissä ovat mukana myös kaupungit [55].

Kunnanimistön avulla tapahtuvan paikkatiedon tallennuksen ongelmana onkin se, että kuntanimistö on paikannimistöstä kaikkein vaihtuvimpia. Vaikka kunnan nimi muutoksen yhteydessä säilyisi samana, sen kattama alue voi muuttua huomattavasti erilaisten aluemuutosten vaikutuksesta. Kuvassa 3 esitetään kartalla Joensuun kaupungin alueen kasvu vuodesta 1948 vuoteen 2015. Joensuuhun on liitetty tuona aikana viisi eri kuntaa ja sen pinta-ala on kasvanut 159-kertaiseksi.



Kuva 3. Joensuu vuonna 2015 tarkoittaa alueellisesti aivan eri asiaa kuin Joensuu tarkoitti vuonna 1948 [56]. Karttapohjat: Maanmittauslaitos.

Vaikka hallinnollisissa alueissa tapahtuneet muutokset vaikuttavat alueiden nimillä tapahtuvien hakujen löydettävyyteen, muutostietoja Suomen kunnista on ollut hankalasti saatavilla. Tarve muutostietojen keräämiseen heräsi kehittäessäni Geologian tutkimuskeskuksen valokuvatietokantaa 2000-luvun alussa [54]. Aloitin silloin muutostietojen järjestelmällisen kokoamisen Geologian tutkimuskeskuksen käyttöön. Tietojen kokoaminen jatkui myöhemmin myös Aalto-yliopiston ja Helsingin yliopiston semanttisen laskennan tutkimusryhmässä (SeCo) Suomen ajallista paikkaontologiaa (SAPO) varten Suomen kulttuurirahaston rahoittamana [25].

Keräämäni aineiston mukaan Suomen kunnissa on tapahtunut yhteensä 1101 muutostapahtumaa, jotka koskevat 1736 kuntaa [55]. Käytännössä muutoksia on tapahtunut lukumääräisesti enemmän, sillä kaikista pienistä aluesiirroista ei ole löytynyt tietoa. Alueellisesti merkittävämpiä ovat kuitenkin kokonaisia kuntia koskevat muutokset. Taulukossa 1 on tilastoa Suomen kuntien nimistössä ja alueissa tapahtuneista muutoksista.

Taulukko 1. Suomen kunnissa tapahtuneiden muutosten lukumäärät muutostyypeittäin ja kuntien lukumäärät, joita muutokset koskevat vuoteen 2016 mennessä [55].

<i>Muutostyyppi</i>	<i>Määrä</i>	<i>Koskee kuntaa</i>
Alueluovutus	66	66
Aluesaaminen	1	1
Aluesiirto	183	363
Aluesiirtojakautuminen	20	71
Aluesiirtoperustaminen	21	69
Aluevuokraus	2	7
Aluevuokrauksen loppu	2	6
Jakaantuminen	101	204
Lakkauttaminen	1	1
Nimenmuutos	37	37
Perustaminen	488	488
Yhdistyminen	179	423

Hallinnollisten alueiden muutoksilla on merkitystä, kun paikkatietoa tallennetaan tietojärjestelmiin tai haetaan tietojärjestelmistä. Esimerkiksi eriaikaisia tilastotietoja verrattaessa on huomioitava, etteivät samalla paikannimellä tallennetut tiedot välttämättä käsitä samaa spatiaalista aluetta [6]. Kuvan 3 perusteella voidaan huomata, että eri vuosina lasketut väestötilastot Joensuun kunnasta eivät ole vertailukelpoisia, ellei oteta huomioon kaupungin alueessa samana aikana tapahtuneita aluemuutoksia. Esimerkiksi kun harvaan asuttu maalaiskunta liitetään kaupunkiin, kaupungin asukastiheys putoaa huomattavasti. Tämä ei kuitenkaan tarkoita sitä, että ihmiset asuisivat alueella yhtään harvemmassa kuin ennenään.

Erilaisissa kokoelmatietokannoissa paikkatieto merkitään luetteloitavan tapahtuman ajankohdan mukaan, jottei alkuperäistä paikkatietoa menetetä. Tietokantaan tallennettuja kohteita haettaessa kaikki vanhat kohteet eivät löydy nykyistä paikannimistöä käyttämällä, jos nimistössä on tapahtunut muutoksia [22, 30]. Esimerkiksi haettaessa kunnan nimellä Outokumpu ei löydetä vanhoja valokuvia Kuusjärveltä, joka on sama kunta ennen nimenmuutosta vuonna 1968. Vanhojen kunnanimien perusteella haettaessa taas eivät löydy aluemuutosten jälkeen samoilta alueilta otetut uudemmat kuvat. On siis olemassa tarve ratkaista tämänkaltaisia ongelmia tietojärjestelmien tiedon tallentamisessa ja löytämisessä.

Tämä työ pyrkii vastaamaan seuraaviin tutkimuskysymyksiin:

1. Millaisia aluemuutoksia Suomen hallinnollisissa alueissa on tapahtunut, ja miten tapahtuneet muutokset vaikuttavat tietojärjestelmiin tallennetun tiedon löydettävyyteen?
2. Mitä erilaisia menetelmiä voidaan käyttää, jotta paikkatiedossa tapahtuvat muutokset tulevat otetuksi huomioon tietoja tietojärjestelmiin tallennettaessa ja niistä haettaessa, ja miten menetelmät vaikuttavat tietojärjestelmiin tehtävien hakujen tarkkuuteen ja saantiin?
3. Miten suomalaisissa muisti- ja tutkimusorganisaatioissa tallennetaan ja haetaan sanallista paikkatietoa ja otetaanko paikkatiedon muuttuminen jotenkin huomioon?

Tutkittavina menetelminä käytetään itse kehittämäni loogisiin operaattoreihin perustuvaa ratkaisua, ajallista paikkaontologiaa SAPOa ja muutostaulua verkkopalvelun käyttämässä valokuvatietokannassa. Hakutulosten vertailu perustuu yleisesti tiedonhakujen toimivuuden arvioinnissa käytettyyn tarkkuus-saanti-analyysiin. Kaikilla tässä työssä esitetyillä käsitteillä ei ole vielä vakiintunutta suomenkielistä nimeä tai osa suomenkielisistä termeistä voi olla monikäsitteisiä. Siksi osaan termeistä on merkitty vakiintuneempi englanninkielinen vastine sulkuihin suomenkielisen termin perään.

Tärkeimpinä lähteinä tässä työssä ovat olleet yhdessä Tomi Kauppisen kanssa kirjoittamani raportti *SAPOn nykytila, ylläpito ja laajennusmahdollisuudet* [56], jossa kuvasimme ensimmäistä kertaa Suomen kunnissa tapahtuneet erityyppisen muutokset, vuonna 2007 kirjoittamani insinööri työ *Ajallisesti muuttuvan paikkatiedon hallinta* [54], jossa ensimmäisen kerran esittelin Geologian tutkimuskeskuksen valokuvatietokannassa toteuttamani menetelmän kuntamuutosten hallintaan sekä *Tomi Kauppisen väitöskirja* artikkeleineen [20], joissa esitellään ja kuvataan Suomen ajallinen paikkaontologia SAPO. Neljässä väitöskirjan sisältämistä artikkeleista olen ollut yhtenä kirjoittajista [20, 22, 23, 24]. Lisäksi olen yhtenä kirjoittajana kirjanluvussa *Representing and utilizing changing historical places as an ontology time series* [10].

Seuraavassa luvussa kaksi perehdytään siihen, millaisia menetelmiä ja tuloksia tieteilisissä julkaisuissa on tämän työn tutkimusaiheesta esitetty. Luvussa kolme esitellään Suomen hallinnollisissa alueissa tapahtuneet muutokset tyypeittäin ja niiden sijoittuminen ajallisesti. Kolmessa seuraavassa luvussa 4–6 esitellään jokaisessa yksi menetelmä, jolla voidaan ratkaista muutosten tiedonhauille tuottamia ongelmia. Luvussa seitsemän tutkitaan menetelmien tuottamaa hyötyä hakutuloksen tarkkuuden ja saannin kannalta sekä esitellään ajalliseen paikkatietoon liittyvän kyselyn tulokset. Luvussa kahdeksan pohditaan työn tuloksia ja aihepiirin tutkimuksen tulevaisuudennäkymiä. Luku yhdeksän kokoaa yhteen diplomityön tärkeimmät havainnot.

2 Taustaa

2.1 Maantieteellinen tiedonhaku

Tiedonhaussa (engl. *Information Retrieval, IR*) kaksi yleisimmin käytettyä mittaria hakujen toimivuudelle ovat tarkkuus ja saanti (engl. *precision and recall*) [32]. Tarkkuus mittaa kuinka suuri osa hakutuloksista on relevantteja ja saanti kuinka suuri osa kaikista relevanteista tuloksista löydettiin [32]. Tarkkuutta ja saantia arvioidessa täytyy tietää, mikä tulos on käyttäjän kannalta relevantti ja mikä ei. Relevanttien kohteiden löytymisen ehtona on taas, että käyttäjän syöttämien hakutermien ja aineiston sisältämien kuvailutietojen välillä on vastaavuuksia. Hakujärjestelmä voi auttaa termien vastaavuuksien löytymisessä esimerkiksi laajentamalla hakua niin, että käyttäjän kannalta relevantit tulokset sisältyvät tulosjoukkoon ilman että epärelevanttien tulosten määrä kasvaa haitallisen suureksi.

Hakumenetelmien toimivuutta arvioidessa tavallista on, että hakutuloksen tarkkuus laskee saannin kasvaessa, koska hakutulosten joukkoon tulee myös epärelevantteja tuloksia. Tässä työssä näiden mittareiden avulla tullaan arvioimaan ajallisen paikkatiedon (engl. *spatio-temporal data*) hallinnassa käytettyjen menetelmien toimivuutta.

Maantieteellisellä tiedonhaulla (engl. *Geographic information retrieval, GIR*) tarkoitetaan tiedonhakua, johon liittyy tavalla tai toisella spatiaalinen ulottuvuus [5]. Nimityksenä GIR on varsin uusi. Ensimmäisenä sen esitti ja nimesi Ray Larson vuonna 1996 [5, 9]. Maantieteellinen tiedonhaku pohjautuu hakijan antamien ja kokoelmassa olevien paikanmääritteiden yhteensovituksen arviointiin. Ne kokoelman kohteet, jotka parhaiten vastaavat hakua, sijoitetaan tulosten relevanssia mittaavien algoritmien avulla hakutulosten kärkeen [9].

Maantieteelliseen informaation tallennukseen ja hakuun liittyy Jonesin mukaan [18] monia erityiskysymyksiä, jotka ovat olleet viime aikoina runsaan tutkimuksen kohteena:

- Paikannimien erottaminen muista käsitteistä luonnollista kieltä olevissa tekstidokumenteissa ja käyttäjien niihin tekemissä kyselyissä
- Epätäsmällisten paikkakäsitteiden maantieteellisen ulottuvuuden määrittäminen sekä niiden käyttäminen informaation tallennuksessa ja haussa (esimerkiksi 'Suomenlahden rannikkoalueella', 'Oulusta itään' tai 'lähellä Helsinkiä')
- Tehokkaan käyttöliittymän kehittäminen spatiaalisia hakuja varten
- Hakujen tulosten osuvuuden arviointi maantieteellisesti ja temaattisesti sekä hakutulosten laittaminen relevanssijärjestykseen
- Kuinka voidaan yksikäsitteistää paikannimillä määritettyjä kohteita, niin että tiedetään mitä tiettyä paikkainstanssia kulloinkin tarkoitetaan.

Osaa näistä ongelmista, varsinkin viimeiseksi mainittua, sivutaan myös tässä työssä, mutta erityisesti keskitytään haasteeseen:

- Kuinka voidaan huomioida paikkakäsitteissä tapahtuvat ajalliset muutokset [16, 21, 34].

Maantieteellisen tiedontallennuksen ja -haun avuksi on kehitetty tapoja hallita paikkatietoa ja paikannimistöä tietojärjestelmissä. Yleisimpiä niistä ovat nimistöluettelot, te-saurukset, ontologiat ja erilaiset GIS-menetelmät. Myös edellä mainittujen yhdistelmiä

on kehitetty parhaan tuloksen saavuttamiseksi. Kehittyneimmissä järjestelmissä esimerkiksi ontologioiden ja GIS-menetelmien yhdistelmillä on saavutettu hyviä tuloksia [40]. Näistä eri menetelmistä kerrotaan tarkemmin seuraavissa kappaleissa.

2.2 Nimistöluettelot

Asiasanoituksen pohjana on käytetty ja käytetään usein edelleen sanaluetteloita, asiasanastoja tai tesauksia, joista poimitaan käytettävät termit. Yksinkertaisimmillaan maantieteelliset sanastot (engl. *gazetteer*) sisältävät vain kolme peruselementtiä, joita ovat paikannimet, paikkatyypit ja sijaintitiedot [9]. Sijaintitieto esitetään yleisimmin koordinaattiparilla, mutta se voi olla myös viiva tai alue.

Monipuolisemmat paikannimistöluettelot voivat olla tesauksia tai ontologioita, jotka sisältävät käsitteiden välisiä suhteita. Nämä suhteet voivat Janowiczin ja Keßlerin [16] mukaan olla esimerkiksi hierarkkisia suhteita, assosiatiivisia suhteita tai vastaavuussuhteita. Hierarkkiset suhteet voivat tarkemmin luokiteltuina olla partitiivisia (osa/kokonaisuus -suhteita), geneerisiä (alaluokkasuhteita) tai instanssisuhteita (yksilön ja sen tyypin välisiä). Tesauksissa hierarkkisia suhteita kuvataan yleensä merkinnöillä *laajempi termi* tai *suppeampi termi*. Niissä on harvoin luokiteltu suhdetyyppejä sen tarkemmin, toisin kuin ontologioissa. Assosiatiivisilla suhteilla kuvataan käsitteiden välisiä yhtäläisyyksiä, jotka eivät ole hierarkkisia. Vastaavuussuhteilla taas esitellään muita termejä, jotka tarkoittavat samaa käsitettä.

Asiasanastoja ja tesauksia kutsutaan kontrolloiduiksi sanastoiksi. Se johtuu siitä, että niissä aina vain yksi termi valitaan edustamaan kutakin käsitettä. Sitä kutsutaan suositeltavaksi termiksi (engl. *preferred term*). Muut samaa käsitettä tarkoittavat termit (synonyymit) merkitään ohjeistuksesta riippuen ei-suositeltaviksi termeiksi (engl. *non-preferred terms*) [16], rinnakkaisiksi termeiksi (engl. *variant terms*) [9] tai sallittaviksi termeiksi (engl. *admitted terms*) [37]. Koska suositeltava termi voi olla kerrallaan vain yhdellä kielellä, asiasanastot ja tesaukset ovat kieliriippuvaisia, koska ne pohjautuvat termeihin eivätkä kieliriippumattomiin yksilöiviin tunnisteisiin [9]. Tästä syystä esimerkiksi Yleisessä suomalaisessa asiasanastossa (YSA) suositellaan käyttämään termiä Kauniainen² samasta paikasta kuin Allmän tesaurus på svenska (Allärs) suosittelee käyttämään termiä Grankulla³. Erikielisiä termejä yhdistämään on tesauksiin voitu luoda vastaavuussuhteet termien välille. Näin on myös tehty YSA:n ja Allärsin tapauksessa.

Paikannimien yksikäsitteistäminen (engl. *disambiguation*) tapahtuu nimistölueteloissa käyttämällä hyväksi itse paikannimen lisäksi muita paikkaan liittyviä tietoja – tyypillisesti paikkatyyppiä ja sijaintitietoja [9]. Paikkatyyppin avulla pystytään erottamaan toisistaan esimerkiksi Pyhäjärven kunta ja sen sisällä sijaitseva järvi, joka on nimeltään myös Pyhäjärvi. Sijaintitiedon avulla voidaan taas erottaa toisistaan eri puolilla Suomea sijaitsevat järvet nimeltään Pyhäjärvi, joita on Kansalaisen karttapaikan mukaan 39 kappaletta [30]. Jotta kohteen yksikäsitteistämisestä olisi hakuja tehtäessä hyötyä, täytyisi sen olla mahdollista sekä tietoja tallennettaessa että haettaessa [20, 50]. Nimistöluetteloiden avulla tätä on hankala toteuttaa, mutta ontologioiden avulla ongelmatonta.

Jotkut maantieteelliset nimistöluettelot kuten Getty Thesaurus of Geographic Names⁴ (TGN) sisältävät eri kielisten nimien lisäksi myös eriaikaisia paikannimiä [27].

² <http://www.yso.fi/onto/ysa/Y94199>

³ <http://www.yso.fi/onto/allars/Y32280>

⁴ <http://www.getty.edu/research/tools/vocabularies/tgn/>

Ilman ajallista ulottuvuutta tesauukset eivät voi esittää paikannimissä, sijaintitiedoissa ja paikkojen välisissä suhteissa ajan kuluessa tapahtuvia muutoksia, vaan pitäytyvät aina viimeisimmässä tilanteessa [9]. Esimerkiksi International Standards Organisationin (ISO TC 211) määrittämässä standardissa *Geographic Information – Spatial Referencing by Geographic Identifiers* (ISO 19112:2003) ainoa paikkoja koskeva ajanmääre, *Temporal extent*, tarkoittaa päiväystä, jolloin paikkainstanssin versio on luotu järjestelmään [14].

Getty Thesaurus of Geographic Names käyttää omaa tesaurusmalliaan paikkatiedon hallinnassa. TGN:n hierarkia perustuu partitiivisiin suhteisiin [9]. Laajempien ja suppeampien termin väliset suhteet ovat siis osa/kokonaisuus -suhteita (engl. *part-of relationships*). Esimerkiksi Helsinki on osa Uuttamaata, joka on osa Suomea, joka on osa Eurooppaa, joka on osa Maailmaa. Kuvassa 4 näytetään Helsinkiin liittyvät tiedot TGN:ssä. Siinä kaikilla paikoilla on yksilöivät ID-tunnisteet ja paikkatiedot esitetään pistekoordinaateilla tai rajauslaatikoilla (engl. bounding box). Helsinki on merkitty hallinnollisen alueen suositeltavaksi nimeksi ja Helsingfors rinnakkaisnimeksi. Nimen kieltä ei ole kuitenkaan määritetty tarkemmin kuin koodilla v, joka tarkoittaa että se on ”paikallisella kielellä”.

TGN sisältää myös historiallisia paikkoja ja niiden olemassaololle on mahdollista antaa alku- ja loppupäiväykset. Suuri osa historiallisesta tiedosta on TGN:ssä kuitenkin vain sanallisessa muodossa. Esimerkiksi Helsingin osalta kerrotaan sanallisesti, että kaupunki perustettiin alun perin Vantaanjoen suistoon ja siirrettiin nykyiselle paikalleen vuonna 1748. Paikkainstanssin tiedot on mahdollista ladata verkkopalvelusta myös semanttisen webin käyttämissä formaateissa RDF/XML ja Turtle.

ID: 7009880

Record Type: **administrative**

 **Helsinki (inhabited place)**

Coordinates:






Lat: 60 08 00 N *degrees minutes* Lat: 60.1333 *decimal degrees*
Long: 025 00 00 E *degrees minutes* Long: 25.0000 *decimal degrees*

Note: Located on peninsula in Gulf of Finland; founded by King Gustav Vasa of Sweden in the 16th century to compete with Reval (Tallinn), Estonia across the gulf; ravaged by plague in 1710, and fires in 1710 and 1808; prospered after being strongly fortified in 1748.

Names:

Helsinki (**preferred**,C,V)
Helsingfors (C,V)

Hierarchical Position:

 World (facet)
 Europe (continent) (P)
 Suomi (nation) (P)
 Uusimaa (region (administrative division)) (P)
 Helsinki (inhabited place) (P)

Place Types:

inhabited place (**preferred**, C) founded in 1550 on the Vantaa estuary, moved to present site S of the estuary in 1640
city (C)
national capital (C)
port (C)
provincial capital (H) of the former province of Uusimaa until 1998

Kuva 4. Helsinki Getty Thesaurus of Geographic Names -paikannimipalvelussa.

Eri aikoina käytössä olleiden paikannimien tarkan käyttöajan määrittäminen voi olla vaikeaa ja osa nimiluetteloiden vanhemmista nimiversioista ei välttämättä ole todellisia yleisessä käytössä olleita nimiä, vaan ne voivat johtua jopa yksittäisistä kirjoitusvirheistä [39]. Kirjoitusvirheet ovat yleisiä varsinkin, jos nimistön kerääjä on ollut erikielinen kuin kerättävä nimistö. Nykyään monet nykyisten ja historiallisten paikkojen nimistöluetteloista, TGN:n lisäksi esimerkiksi The Historical Gazetteer of England's Place-

Names⁵ ja Pleiades⁶, löytyvät internetistä ja voivat näyttää nimien ja paikkatyyppien lisäksi haetun paikan sijainnin kartalla.

Erikoistuneissa paikannimistöluetteloissa, jotka sisältävät vanhoja paikannimiä, on tärkeää erottaa vanhat nimet nykyisistä. Siksi paikat luokitellaan esimerkiksi *entisiin* ja *nykyisiin* [9]. Joskus on tarpeen ottaa myös tulevaisuus huomioon, jolloin luokaksi on valittavissa *ehdotetut* (engl. *proposed*), johon voidaan lisätä suunnitellut, mutta vielä toteutumattomat paikat. Vielä tarkempaa olisi ilmoittaa paikan olemassaolon ajanjakso alku- ja päättymispäiväysten avulla. Tämä on otettu huomioon esimerkiksi ADL Gazetteer Content Standardissa (GCS), jonka mukaan alku- ja loppuaikamääreet voidaan antaa paitsi paikannimille, myös paikkojen sijainti-, tyyppi- ja suhdetiedoille [9].

Yleisessä Suomalaisessa Asiasanastossa⁷ (YSA) paikkakäsitteiden ajallisia muutoksia hallitaan samalla tavalla kuin synonyymejä, käyttämällä rinnakkaistermejä tai assosiativisia suhteita. Esimerkiksi, kun Maaningan kunta liitettiin Kuopion kaupunkiin 1.1.2015, YSAan Kuopio merkittiin Maaningan assosiativiseksi käsitteeksi (Kuva 5).

Asiasanaston huomautuskentässä kerrotaan muuttuneiden paikkojen väliset suhteet, esimerkiksi alueliitokset tai jakaantumiset. Valitettavasti tietojärjestelmät eivät voi ilman erityistä ohjelmointia suoraan hyödyntää näitä ihmisluettavia tietoja, vaan ne on manuaalisesti otettava erikseen huomioon kuvailua tai hakua tehtäessä.

KÄYTETTÄVÄ TERMI

Maaninka

TYYPPI

Maantieteellinen paikka

ASSOSIATIIVISET KÄSITTEET

Kuopio

HUOMAUTUS

Maaninka liitetty 1.1.2015 Kuopioon.

LUONTIPÄIVÄ

30.8.1993

MUOKKAUSPÄIVÄ

2.2.2015

URI

<http://www.yso.fi/onto/ysa/Y107738>

Lataa tämä käsite:

[RDF/XML TURTLE](#)

Kuva 5. Maaningan kunta Finto-palvelun Yleisessä suomalaisessa asiasanastossa⁸

Hakutulosten parantamiseksi on monissa tietokannoissa ollut käytäntönä lisätä aineistoja kuvailtaessa luetteloitavan tapahtuman ajankohtaisten paikannimien lisäksi mahdollinen myöhempi kohdetta koskeva paikannimi sulkuihin alkuperäisen paikannimen perään, esim. Siam (nyk. Thaimaa). Näin suositellaan toimimaan esimerkiksi museotietojärjestelmä Musketin luettelointiohjeessa muuttuneiden valtionimien ollessa kyseessä [48]. Tällä tavalla saadaan kohde mukaan hakutulokseen, haettiinpa kummalla nimellä tahansa. Ongelmana tässä tavassa on se, että kaikki muuttuneeseen paikkaan viittaavat aiemmatkin tietueet on uusittava muutoksen tapahduttua. Tämä voi vaatia suurissa tietokannoissa isoja päivityksiä. Alkuperäisiä tietoja huolimattomasti muokattaessa voidaan virheellisellä päivityksellä aiheuttaa suurta vahinkoa.

Helpommin ylläpidettävä tapa on hallita muuttuneita tietoja tietokantaan lisätyllä erillisellä muutostaululla, joka sisältää paikannimien eri ajalliset versiot. Tietojen muuttuessa riittää pelkkä muutostaulun päivittäminen eikä alkuperäisiin tietueisiin tarvitse

⁵ <http://www.placenames.org.uk/>

⁶ <http://pleiades.stoa.org/>

⁷ <https://finto.fi/ysa/fi/>

⁸ <https://finto.fi/ysa/fi/page/Y107738>

tehdä muutoksia. Tätä menetelmää käytettiin The Great Britain Historical GIS -projektin (GBHGIS) ensimmäisessä vaiheessa [6].

Puutteena nimistöluetteloissa on se, että vaikka ne sisältävät historiallisia nimiä ja niiden ajallisia ulottuvuuksiakin, ne eivät useinkaan kuvaa, millaisia muutoksia paikkojen välillä on tapahtunut ja miten muutokset vaikuttavat paikkojen välisiin suhteisiin [20]. Näihin kysymyksiin tuovat apua ontologiat, joita käsitellään seuraavassa kappaleessa. Paikoissa tapahtuneita muutoksia kuvaamaan on kehitetty erityisiä ajallisia paikkaontologioita (engl. *spatio-temporal ontologies*).

2.3 Ontologiat

Ontologialla tarkoitetaan informaatiotekniikassa mallia, joka formaalilla tavalla kuvaa tiettyyn aihealueeseen liittyvää jaettua käsitteistöä [8, 44]. Määritelmässä formaali tarkoittaa sitä, että ontologian on oltava määrämuotoisessa koneluettavassa muodossa. Jaettu puolestaan tarkoittaa, että käsitteistö on tarkoitettu laajempaan kuin yksityiseen käyttöön [46].

Janowich ja Keßler ovat esittäneet (taulukko 2) keskeisimpiä tesarusten ja ontologioiden välisiä eroja [16]. Molempia käytetään nykyään yleisesti apuna tiedon kuvailussa ja haussa, mutta ontologiat ovat tarkemmin määriteltyjä. Ontologioissa suhteiden määrä on vapaa, mutta tesaruksissa käytössä on rajallinen määrä erilaisia suhteita, tyypillisesti ne ovat maantieteellisissä tesaruksissa koostumussuhteita (Suomi - Helsinki), mutta myös hierarkiasuhteita (esimerkiksi kaupungit - pääkaupungit) tai luokka-instanssisuhteita (kaupungit - Kuopio) voidaan käyttää. Ontologiassa voidaan käyttää rajattomasti erilaisia suhteita, hierarkia- ja koostumussuhteiden lisäksi esimerkiksi naapurussuhteita ja peittävyysuhteita, kunhan suhteet on määritelty tarkasti.

Taulukko 2. Tesarusten ja ontologioiden vertailu Janowichin ja Keßlerin mukaan [16]

Näkökulma	Tesarus	Ontologia
Tarkoitus	Tiedon haku ja jäsentäminen	Tiedon haku ja päättely
Rakenne	Hierarkiasuhde, koostumussuhde tai luokka-instanssi-suhde	Hierarkiasuhde ("on")
Suhteet	Rajallinen määrä suhteita	Suhteiden määrä on vapaa
Mitä esitetään	Termit edustavat käsitteitä	Käsitteet ja niiden suhteet määritellään tarkasti
Semantiikka	Ei määrämuotoista semantiikkaa	Määrämuotoinen semantiikka

Ontologioissa käsitteille annetaan Uniform Resource Identifier- eli URI-tunnisteet [52] (esimerkiksi http://www.ys.fi/onto/sapo/Pyhajarvi_1969_01_01). Sen ansiosta samannimiset paikat voidaan aina yksilöidä. Oikean paikkainstanssin (ja sen URIn) valintaa auttaa, kun ontologia sisältää runsaasti sijaintiin ja aikaan liittyviä ominaisuustietoja [21]. Ontologiat ovat periaatteessa myös kieliriippumattomia, koska käsitteille voidaan aina määritellä nimet useilla eri kielillä ominaisuustietoihin, vaikka alkuperäinen ontologia olisi laadittu vain yhdellä kielellä. Esimerkiksi SAPO-ontologiassa käsitteellä

http://www.yso.fi/onto/sapo/Kauniainen_1949_04_08 on sekä suomenkielinen nimi Kauniainen (1949-04-08) että ruotsinkielinen nimi Grankulla (1949-04-08).

Aikaan ja paikkaan liittyvän päättelyn mahdollisuus on yksi syy käyttää ontologioita paikkatietojen hallinnassa tesausten sijasta [44]. Naapuruussuhdetta voidaan käyttää esimerkiksi spatiaalisen haun tuloksen relevanssia nostavana tekijänä [45]. Voidaan esimerkiksi laajentaa haku ”Helsingin seudulla” kattamaan Helsingin naapurikunnat Espoo, Vantaa ja Sipoo.

Perinteisistä tietokannoista tietoa haetaan asiasanoilla. Ontologioita käytettäessä kuvailu ja haku tehdään asiasanojen sijasta käsitteillä [22]. Kauppinen ja Hyvösen mukaan tämän ansiosta hakutulokseen saadaan enemmän relevantteja osumia eli haun tarkkuus paranee. Lisäksi kaikista relevanteista osumista saadaan tulokseen suurempi osuus eli myös saanti paranee. Tämän lisäksi ontologiat voivat auttaa hakijaa muotoilemaan hakuja paremmin, jos palvelu hyödyntää semanttista päättelyä, esimerkiksi tarjoamalla hakijalle valmiita vaihtoehtoja.

Ontologiat eivät vain tarjoa tehokkaan keinon määritellä aihealueen keskeisiä käsitteitä vaan mahdollistavat erilaisten tietolähteiden yhdistämisen. Tämä perustuu siihen että ontologioissa käsitteet ovat semanttisesti tarkkaan määriteltyjä. Tämä taas mahdollistaa erilaisten semanttisten verkkopalveluiden toteuttamisen [33].

Maantieteellisiin ontologioihin voidaan sisällyttää hyvinkin runsaasti tietoa kohteista. Jones ym. [17] suosittelevat hakuja ajatellen ontologioiden sisältävän esimerkiksi seuraavia tietoja:

- Paikan nimi, rinnakkaisnimet ja nimet eri kielillä
- Sijaintitiedot (edustava piste, rajauslaatikko tai polygoni)
- Osoitetiedot (postinumero ja katuosoite)
- Spatiaaliset suhteet (olisi toivottavaa, että myös käytettyjen suhteiden ontologia olisi määritelty järjestelmässä)
- Käytetyt koordinaattijärjestelmät
- Aikamääreet (perustamis-, lakkauttamis-, nimenmuutos- ja rajan muuttumisajat)

Ontologiaa luotaessa olisi harkittava paikkojen esittämiselle sopiva semanttinen ja geometrinen yleistämisaste eli sisällön ja sijaintitietojen sopiva esitystarkkuus, esimerkiksi kuvataanko paikkaa pisteellä, rajauslaatikolla vai polygonilla [17]. Aikamääreiden sisällyttäminen ontologiaan mahdollistaa sen, että ontologiaan voidaan sisällyttää myös historiallisia paikkoja, esimerkiksi Kulosaaren huvilakaupunki, joka oli olemassa vuosina 1922–1945. Silloin voidaan puhua ajallisesta paikkaontologiasta.

Teoreettisesti tämä vastaa Grenon & Smithin [7] kuvausta Basic Formal Ontologyn (BFO) SNAP-osasta. Ontologia sisältää tällöin vain *olevaiset* (engl. *continuants*) eli yksiköt, jotka säilyttävät identiteettinsä ajan läpi, vaikka läpikäyvätkin erilaisia muutoksia. Grenonin ja Smithin mielestä kokonaisuuden kattamiseksi tärkeää on ottaa huomioon myös SPAN-osa, joka sisältää *tapahtuvaiset* (engl. *occurents*) eli erilaiset prosessit, toiminnot ja muutokset. Käytännössä he tätä kautta suosittelevat ottamaan huomioon esimerkissämme myös mistä tapahtumasta Kulosaaren huvilakaupunki sai alkunsa ja mitä sille tapahtui vuoden 1946 vaihteessa.

2.4 GIS-menetelmät

GIS (Geographic Information Systems) eli paikkatietojärjestelmät auttavat käsittelemään paikkojen sijainti- ja ominaisuustietoja visuaalisesti kartta- tai ilmakuvapohjan avulla [53]. Erilaisten kohteiden sijaintitietoja kuvataan GIS-järjestelmissä koordinaat-

tiparien tai -sarjojen avulla. Pistemäisiä kohteita, kuten suuria kiviä, kuvataan pisteillä (koordinaattipareilla), viivamaisia kohteita kuten sähkölinjoja viivoilla (koordinaattisarjoilla) ja erilaisia alueita polygoneilla (sulkeutuvilla koordinaattisarjoilla) [1].

Hallinnollisia alueita kuvataan polygoneilla, mikäli rajan riittävän tarkka sijainti on tiedossa. Muussa tapauksessa voidaan käyttää myös alueella – esimerkiksi kunnan taa-jamassa – sijaitsevaa pistettä, mutta se ei anna mitään kuvaa alueen mittasuhteista. On myös tutkittu epätarkkojen historiallisten paikkatietojen kuvaamista määrittämällä todennäköisyyksiä, joilla kysytty kohde on sijainnut kartalla [53]. Lopputuloksena kartalla voidaan näyttää tummimmalla värillä todennäköisimmät alueet, jolla kohde on sijainnut ja sen ympäristössä asteittain vaalenevina vyöhykkeinä mahdolliset, mutta vähemmän todennäköiset alueet.

Kun halutaan hallita paikkatiedon muutoksia, nimistön lisäksi toinen lähestymistapa on aloittaa sijaintitietojen muutoksista, esimerkiksi kuntien rajapolygoneissa tapahtuvista muutoksista kahden kunnan liittyessä yhdeksi kunnaksi. Gregoryn ja muiden mukaan muuttuvien sijaintitietojen hallinnassa GIS-menetelmillä avainkysymys on kuinka kytkeä rajapolygonien muutokset GIS-tietomalliin [6]. Ongelmaksi voivat muodostua paikkatieto-ohjelmistojen puutteet aikamääritteiden hallinnassa ilman lisäohjelmointia [41].

Paikannimistöjen hallintaan GIS-menetelmätkin käyttävät tyypillisesti nimistöluette-loita [16]. Nimistöjä hallitaan GIS-ohjelmien attribuuttitietokannoissa, joista ne kytke-tään avainkenttien kautta koordinaattitietoihin. Useiden nimiversioiden hallintaan tarvi-taan oma tietokanta, joka ohjaa eri nimillä tapahtuvat haut GIS-ohjelman käyttämään standardinimeen [6]. Vielä joustavampi ratkaisu on käyttää kytkentään yksilöivää tun-nistetta nimen sijasta, jolloin se on täysin kieliriippumaton eikä ota kantaa siihen, mikä on paikan ”virallinen” nimi. Tähän ratkaisuun päädyttiin esimerkiksi The Great Britain Historical GIS -projektin (GBHGIS) toisessa vaiheessa [39].

Tavallinen tapa kuvata historiallisia alueita GIS-menetelmillä on käyttää tilanneku-via (engl. *snapshots*) halutuilta ajanhetkiltä [4]. Esimerkiksi tilastotietoja on verrattu tilaston tekohetken rajatietoihin. Tilannekuvatietokannat ovat helppoja luoda ja ylläpitää GIS-ohjelmistojen normaalien työkalujen avulla, mutta ne sisältävät runsaasti päällekkäistä tietoa ja epäonnistuvat muutosten kuvaamisessa [6].

Ison-Britannian GBHGIS-projektissa on pyritty luomaan kattava sarja rajatiedoista kautta historian [6]. Projektin ensimmäinen vaihe toteutettiin vuosina 1994–1999 ja jälkimmäinen vaihe jatkuu edelleen. Projektissa on keskitytty ennen kaikkea historialli-sen tilastotiedon, esimerkiksi väestönlaskennan tietojen, kytkemiseen hallinnollisten alueiden paikkatietoon.

Lähtömateriaalina projektissa käytettiin saatavilla olevia eriaikaisia karttoja rajatie-toineen, kun taas aluemuutosten tarkoista tapahtuma-ajoista saatiin tietoja kirjallisista lähteistä [6]. Eriaikaiset rajatiedot digitoitiin ArcGIS-järjestelmään irrallisina viivoina (arc), jotka voitiin yhdistää polygoneiksi erillisellä aputyökalulla. Sen ansiosta koko-naisten polygonien uudelleenpiirtämiseltä rajojen muuttuessa vältyttiin. Riitti, että päi-vitettiin vain muuttunut osa rajasta. Tärkeimmiksi GBHGIS-projektissa vuosien 1840–1974 väliseltä ajalta havaituiksi muutosmuodoiksi Gregory ja muut luokittelevat [6] aluesiirrot (transfers), nimenmuutokset (name changes), yhdistymiset (mergers) ja ja-kaantumiset (divisions). Muut tapahtuneet muutokset pystyttiin mallintamaan edellisten avulla. Alueiden eriaikaiset ja rinnakkaiset nimet hallittiin projektissa niin, että paikka-tietojärjestelmässä käytettyyn yhteen valikoituun nimeen kytkettiin nimistötietokanta, jossa hallittiin vaihtoehtoisia nimiä.

Ison-Britannian historiallisen paikkatietoprojektin kattaman alueen laajetessa ja ajas-sa taaksepäin mentäessä huomattiin, että kaikkien hallinnollisten alueiden polygoneja ei pystytty välttämättä koskaan määrittämään [39]. Siksi GIS-keskeisestä järjestelmästä

siirryttiin käyttämään viitetietokantaa tietojen tallentamiseen. Kytkeä paikkatietoon säilytettiin tietokantaan tallennettujen paikkatunnisteiden avulla. Paikkakäsitteistä muodostettiin Administrative Unit Ontology (AUO) -ontologia, jota hallitaan tietokannan avulla.

Erityyppisiä rajatietoja mallintavia kansallisia GIS-projekteja on toteutettu Ison-Britannian lisäksi monissa muissakin maissa, mm. Saksassa, Belgiassa, Hollannissa, Ruotsissa ja USA:ssa [6]. Niiden toteutukset ovat vaihdelleet laajuudeltaan, kustannuksiltaan ja ajalliselta kattavuudeltaan huomattavasti. Yksinkertaisimmillaan projekti on linkittänyt tietyiltä valikoiduilta vuosilta olevaa aineistoa samojen vuosien hallinnollisten alueiden rajatietoihin, kuten projektissa Historical Atlas of Prussia [51]. Monipuolisemmissa projekteissa, kuten aiemmin mainitussa Ison-Britannian GBHGIS:ssä, on onnistuttu toteuttamaan hallinnollisten alueiden rajoista kattavia paikkatietokantoja pidemmille aikaväleille.

GIS-menetelmien suurimmaksi ongelmaksi GBHGIS-projektissa mukana olleet Gregory ja muut esittävät GIS-tietokantojen kehittämisen vievän runsaasti aikaa, olevan huomattavan kallista ja vaativan monimutkaista tietoarkkitehtuuria [6]. Pelkkiä tilannekuvia monipuolisempi mallintaminen vaatii myös GIS-osaamista ja toteutuksen tavasta riippuen mahdollisesti myös räätälöityjen lisäohjelmien ohjelmointia. Esimerkiksi GBHGIS-projektin toteutus ei olisi ollut mahdollista pelkästään normaaleilla ArcGIS-ohjelman tarjoamilla työkaluilla.

3 Muuttuva paikkatieto Suomessa

Tässä työssä pohjamateriaalina ovat Suomen hallinnollisissa alueissa tapahtuneista muutoksista keräämäni tiedot. Muutostietojen kerääminen alkoi Geologian tutkimuskeskuksen valokuvatietokantaa varten ja jatkui myöhemmin myös Suomen kulttuurirahaston rahoittamana Suomen ajallista paikkaontologiaa (SAPO) varten. Tämän kappaleen ensimmäisessä luvussa kerron millaisia muutoksia Suomen hallinnollisissa alueissa on ajan kuluessa tapahtunut ja miten ne ovat vaikuttaneet paikkojen ominaisuuksiin, esimerkiksi nimiin ja pinta-aloihin. Toisessa luvussa esitän, miten Suomen kunnissa tapahtuneet muutokset sijoittuvat aikajanelle ja pohdin muutoshistorian vaikutusta paikannimistön mukaan tehtyihin hakuihin. Koska pinta-alatieto on yksi muutosten mallintamisessa käytetty ominaisuus, esitän kolmannessa luvussa syitä, miksi hallinnollisen alueen pinta-alan mittaustieto voi muuttua ilman päätöstä alueen rajojen muuttamisesta.

3.1 Suomen hallinnollisissa alueissa tapahtuneet muutokset

Hallinnollisia alueita ovat esimerkiksi valtiot, osavaltiot, läänit, maakunnat, kihlakunnat, seutukunnat ja kunnat. Hallinnollisten alueiden jako on tyypillisesti hierarkkinen. Esimerkiksi kunnat ovat aina historian kuluessa kuuluneet suurempiin hallinnollisiin alueisiin kuten lääneihin tai maakuntiin. Taulukossa 3 esitetään Suomen alueella sijainneita hallinnollisia alueita ajallisine ulottuvuuksineen. Tässä työssä hallinnollisista alueista keskitytään ennen kaikkea kuntiin, mutta samoja periaatteita voidaan soveltaa myös muihin hallinnollisiin alueisiin.

Taulukko 3 Hallinnollisten alueiden tyyppejä Suomen alueella eri aikoina

Aluehallintovirastot	2010	–
ELY-keskukset	2010	–
Hallintopitäjät	–	1865
Kihlakunnat	–	2007
Kirkkopitäjät	–	1892
Kunnat	1865	–
Läänit	1285	– 1997
Maakunnat (toiminnallis-taloudelliset)	1994	–
Valtiot	1250	–

Hallinnollinen aluejako ei ole ajan suhteen pysyvä, vaan heijastaa aina kulloistakin hallintoa. Olen tutkinut Suomen alueella vuodesta 1865 lähtien tapahtuneita hallinnollisten alueiden muutoksia ja havainnut kahdeksan perusmuutostyyppiä. Kuusi niistä on Suomen valtion sisäisiä ja kaksi kahden valtion välisten aluesiirtojen aiheuttamia. Uusiina muutostyyppinä olen havainnut kaksi perusmuutostyyppien yhdistelmää, jotka on hyvä niiden monimutkaisuuden vuoksi käsitellä omina muutostyyppinä. Tämän työn käsittämänä aikana on tapahtunut lisäksi rauhansopimuksiin liittyviä aluevuokrauksia Suomesta Neuvostoliitolle, jotka voidaan lakkaamisineen myös laskea omiksi muutostyypeikseen. Jos menemme ajallisesti vielä historiassa paljon kauemmaksi, joudumme

ottamaan käyttöön vielä kaksi muutostyyppiä, jotka esittelen kappaleen lopussa. Luettelimestani Suomen 14:sta muutostyyppistä kahdeksan kuvattiin ensimmäisen kerran julkaisussa *Modeling and reasoning about changes in ontology time series* [22] ja insinööriyössäni *Ajallisesti muuttuvan paikkatiedon hallinta* [54] vuonna 2007. Loput kuusi esittelimme tänä vuonna raportissa *SAPOn nykytila, ylläpito ja laajennusmahdollisuudet* [56], joka on ollut myös tämän kappaleen lähteenä.

Nimenmuutosta lukuun ottamatta kaikissa muutostyypeissä tapahtuu muutoksia myös alueiden pinta-aloissa. Kuten Kauppinen ja Hyvönen esittävät [22], pinta-aloja voidaan käyttää hyväksi muuttuneita alueita koskevien aineistojen sijaintitodennäköisyyksien laskemisessa. Siihen liittyen esitän jokaisen muutostyyppin osalta graafisesti, miten sijaintitodennäköisyydet eri alueilla voidaan pinta-alojen perusteella laskea. Osasta muutostyyppistä näytän myös esimerkin tapahtuneesta kuntamuutoksesta karttapolygonien avulla.

Kaikista muutostyypeistä käytetyt englanninkieliset nimet eivät ole vielä täysin vakiintuneet. Olen merkinnyt muutoksen suomenkielisen nimen perään eri lähteissä käytettyjä muutosten englanninkielisiä nimiä lähteineen, jotta vertailu kirjallisuudessa esitettyihin muutoksiin olisi helpompaa. Osaa kuvaamistani muutostyypeistä ei ole mainittu missään käyttämässäni lähteessä, joten niistä puuttuvat myös englanninkieliset nimet.

Perustaminen (engl. establishment [25])

Suurin osa Suomen kaupungeista ja kunnista on perustettu ennen Suomen valtion itsenäistymistä. Kaupungin perustamisoikeus oli Ruotsin ja Venäjän vallan aikana vain hallitsijalla. Itsenäisyyden aikana kaupungin tai kauppalan perustaminen vaati senaatin päätöksen. Nykyään mikä hyvänsä kunta voi julistautua kaupungiksi.

Kuntia alettiin Suomeen perustaa 6. helmikuuta vuonna 1865 annetun kunnalliseasetuksen mukaisesti. Asetuksen mukaan kunnat perustettiin olemassa olevien seurakuntien eli kirkkopitäjien alueelle [38]. Anomuksesta kunta saatettiin perustaa myös kappeliseurakuntaan. Kunnan perustamisajaksi tässä työssä katsotaan ensimmäisen kuntakokouksen pitohetki. Perustamiseksi ei tässä työssä lasketa sitä, jos kunta muodostetaan jo olemassa olevien kuntien alueita jakamalla tai yhdistämällä, vaan silloin puhutaan jakaantumisesta tai aluesiirtoperustamisesta. Myöskään kuntamuodon muutosta kunnasta kaupungiksi vuoden 1976 kunnallislain voimaantulon (1.1.1977) jälkeen ei katsota perustamiseksi. Kuva 6 esittää Paikka 1 nimisen paikan perustamista pinta-alan kannalta.



Kuva 6. Hallinnollisen alueen Paikka1 perustaminen. Alueelle, jossa ei ole aiemmin ollut vastaavantyyppistä hallinnollista aluetta, perustetaan sellainen. Perustettavan hallinnollisen alueen pinta-ala on A.

Lakkauttaminen

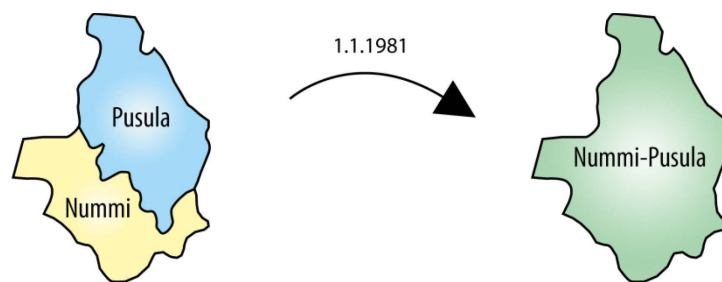
Hallinnollinen alue katsotaan lakkautetuksi, jos se päätetään lakkauttaa ja sen alueelle ei tule enää uutta vastaavantyyppistä hallinnollista aluetta. Merkittävä hallinnollisten alueiden lakkauttaminen tapahtui Suomessa vuoden 2010 alussa, jolloin läänit lakkautettiin. Kuvassa 7 on Paikka 1:n lakkauttaminen pinta-alan kannalta esitettyinä.



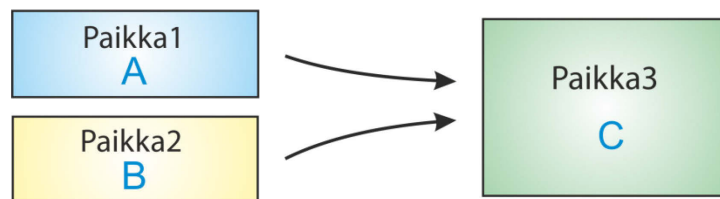
Kuva 7. Hallinnollinen alue Paikka1 lakkautetaan. Lakkautettavan hallinnollisen alueen pinta-ala on A .

Yhdistyminen (engl. *merge* [22], *merger* [4, 6])

Hallinnollisten alueiden yhdistyminen tapahtuu kun kaksi tai useampi hallinnollinen alue yhdistyvät yhdeksi hallinnolliseksi alueeksi. Yhdistyneelle alueelle annetaan yleensä jonkin liitoksessa mukana olevan alueen nimi. Näin on tehty yli 90 %:ssa Suomen kuntaliitoksista [55]. Toisin tehtiin vuoden 1981 alussa, kun Nummi ja Pusula yhdistyivät Nummi-Pusulaksi. Kuvassa 8 yhdistyminen esitetään kuntien rajapolygonien avulla. Lähes aina yhdistyvät alueet ovat olleet rajanaapureita. Jollei yhteistä rajaa muuten ole ollut, on yhteys joskus luotu tekemällä käytävä kolmannen kunnan kautta. Vuoden 1997 kuntajakolain mukaan kuntaliitos ilman yhteistä rajaa ei ole enää mahdollinenkaan. Kuvassa 9 esitetään graafisesti, miten yhdistyneen hallinnollisen alueen pinta-ala muodostuu yhdistyvien alueiden pinta-alojen summasta.



Kuva 8. Nummen ja Pusulan kuntien yhdistyminen Nummi-Pusulaksi. [54]

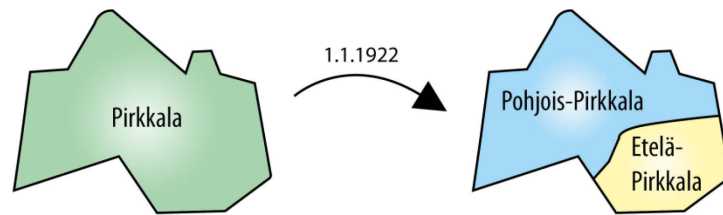


Kuva 9. Hallinnolliset alueet Paikka1 ja Paikka2 yhdistyvät Paikka3:ksi. Pinta-ala $C = A + B$. Esimerkiksi Paikka3:ssa sijaitsevan pisteen todennäköisyys sijaita Paikka1:ssä on $A / (A + B)$. Paikka1:ssä sijainneen pisteen todennäköisyys sijaita Paikka3:ssa on 1.

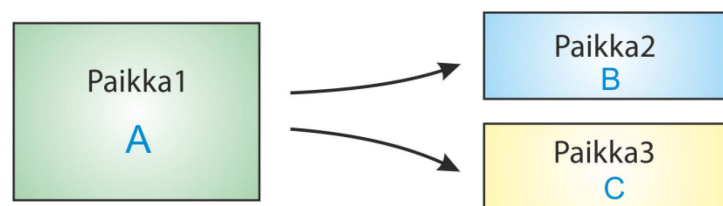
Jakaantuminen (engl. *division* [6], *separation* [4], *split* [22])

Hallinnollisen alueen koko alue jaetaan kahteen tai useampaan uuteen hallinnolliseen alueeseen. Yhtenä syynä tähän voi olla kulkuyhteyksien hankaluus kunnan eri osista päätaajamaan, kuten kuvan 10 Pirkkalan jakaantumisessa. Suurin osa jakaantumisista Suomessa on kuitenkin tapahtunut, kun maalaiskunnan taajama-alueelle on perustettu kauppala tai kaupunki. Maalaiskunnat ovat tällöin jääneet taajamaa ympäröiviksi “reikä-

leipäkunniksi”. Kuvassa 11 esitetään hallinnollisen alueen jakaantumisessa tapahtuvat pinta-alamuutokset.



Kuva 10. Tampereen länsipuolella sijainnut Pirkkalan kunta jakaantui 1.1.1922 Pyhäjärven kohdalta Pohjois- ja Etelä-Pirkkalaksi. [54]



Kuva 11. Hallinnollinen alue Paikka1 jakaantuu kahteen hallinnolliseen alueeseen Paikka2 ja Paikka3. Pinta-ala $A = B + C$. Yleisesti: syntyvien hallinnollisten alueiden pinta-alojen summa on sama kuin jakaantuvan alueen pinta-ala. Esimerkiksi Paikka1:ssä sijaitsevan satunnaisesti valitun pisteen todennäköisyys sijaita Paikka2:ssa on $B / (B + C)$.

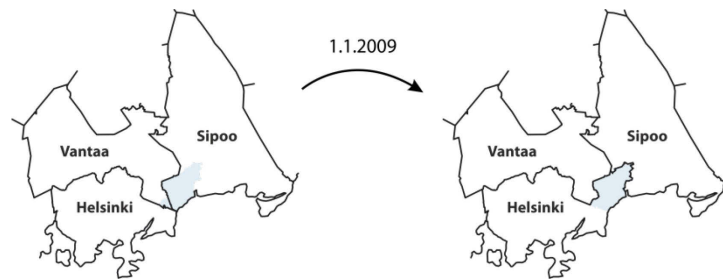
Aluesiirto (engl. transfer [6], land exchange [4], changepart [22])

Kunnista on siirretty alueita toisiin kuntiin monista eri syistä: kasvava taajama on tarvinnut lisää maata ympäröivästä maalaiskunnasta, alueelta on voinut olla paremmat kulkuyhteydet naapuri- kuin oman kunnan palveluihin tai toisen kunnan aluetta oman kunnan sisällä oleva enklaavi on haluttu liittää ympäröivään kuntaan. Läänien väliset aluesiirrot koskevat kokonaisia kuntia. Sama kunta ei voi sijaita yhtä aikaa useassa läänissä. Valtioiden välisille aluesiirroille on omat muutostyypit aluesaaaminen ja alue-luovutus. Tämä siksi että valtioiden välisessä aluesiirrossa paikan luonne hallinnollisena alueena muuttuu, koska eri valtioissa hallinnollisille alueille on erilaiset tehtävät.

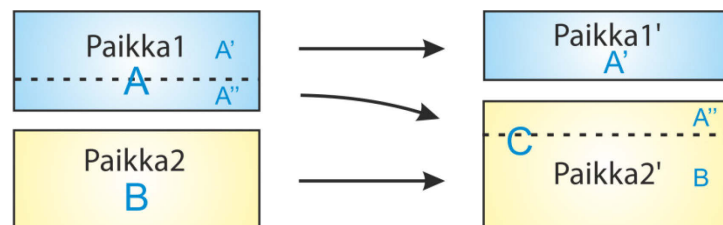
Aluesiirtoja on tapahtunut aikojen kuluessa määrällisesti hyvin runsaasti, mutta pinta-alallisesti ne eivät koske yleensä merkittävän suuria alueita. Pinta-alan suuruus ei tosin ole välttämättä täysin kuvaava mitta siirron yleiselle merkitykselle alueen kunnille ja asukkaille. Vuoden 2009 alussa Sipoo liitettiin Helsinkiin noin 30 neliökilometrin suuruinen alue. Jotta alueeseen saatiin maayhteys, myös Vantaasta liitettiin Helsinkiin ns. Västerkullan kiila. Samalla katkesi Vantaan maayhteys merelle (Kuva 12).

Kuvassa 13 näytetään aluesiirron vaikutukset siirrossa mukana olevien hallinnollisten alueiden pinta-aloihin ja lasketaan siirrossa mukana olevilla alueella sijainneiden kohteiden sijaintitodennäköisyydet siirron jälkeen. Käytännössä vanhemmissa aluesiirroissa siirtyneiden alueiden pinta-alojen suuruuksia on työlästä selvittää. Yleensä on olemassa tieto vain siirtyneen alueen nimestä, mutta ei sen pinta-alasta. Pinta-alatilastoissakaan aluesiirrot eivät aina näy muutoksena hallinnollisen alueen pinta-

alatiedossa. Tämä johtuu puutteellisuudesta pinta-alojen mittaustarkkuudessa ja siitä, ettei vanhemmissa tilastoissa todellisia pinta-aloja ole välttämättä edes mitattu vuosittain.



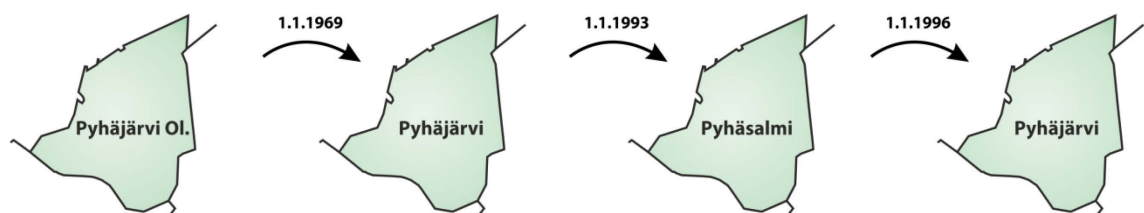
Kuva 12. Vuoden 2009 alussa Sipoon lounaisosasta siirrettiin alue Helsinkiin. Samalla Vantaasta siirrettiin Helsinkiin kiilanmuotoinen maa-alue.



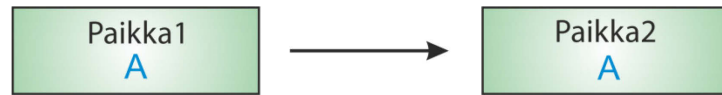
Kuva 13. Hallinnollisesta alueesta Paikka1 siirretään alue pinta-alaltaan A'' hallinnolliseen alueeseen Paikka2. Näin syntyneen Paikka1':n pinta-ala on $A' = A - A''$ ja Paikka2':n pinta-ala $C = A'' + B$. Esimerkiksi Paikka1:ssä sijaitsevan pisteen todennäköisyys sijaita Paikka1':ssa on A'/A ja Paikka2':ssa A''/A .

Nimenmuutos (engl. name change [6, 21])

Nimenmuutokseksi lasketaan hallinnollisen alueen päätös vaihtaa nimeään ilman muuta samanaikaista muutosta. Nimellä on suuri vaikutus hallinnollista aluetta koskevan aineiston löytämiselle. Siksi on tärkeää tietää alueella eri aikoina käytössä olleet nimet. Muutosketju kuvassa 14 esittää Oulun läänissä sijainneen Pyhäjärven kunnan nimen muutokset. Pyhäjärvi-nimisiä kuntia perustettiin Suomeen alun perin kolme. Ne erotettiin toisistaan lääniä kuvaavilla lyhenteillä: Ol., Vpl. ja Ul. Näistä Viipurin läänissä sijainnut Pyhäjärvi jäi luovutetuille alueille ja Uudenmaan läänissä sijainnut Pyhäjärvi liitettiin Karkkilaan. Nimenmuutos ei vaikuta alueen pinta-alaan (Kuva 15).



Kuva 14. Nykyinen Pyhäjärven kunta on vaihtanut nimeään vuosien kuluessa kolmesti.



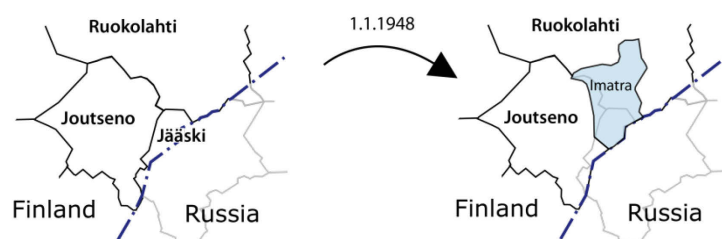
Kuva 15. Nimenmuutoksessa Paikka1:n nimi muuttuu Paikka2:ksi. Alueen pinta-ala pysyy samana.

Nimenmuutokseksi katsotaan vain virallisen päätöksen mukainen nimen vaihtaminen. Monilla kunnilla on ollut varsinkin kauempana menneisyydessä useita kutsunimiä tai erilaisia tapoja kirjoittaa nimi (esimerkiksi Lapvesi, Lappvesi tai Lappee). Näitä ei lasketa nimenmuutoksiksi, vaan merkitään rinnakkaisnimiksi. Nimenmuutoksiksi ei myöskään lasketa, jos kunnan virallinen nimi on muuttunut kunnan kielisuhteiden johdosta ruotsin- ja suomenkielen välillä.

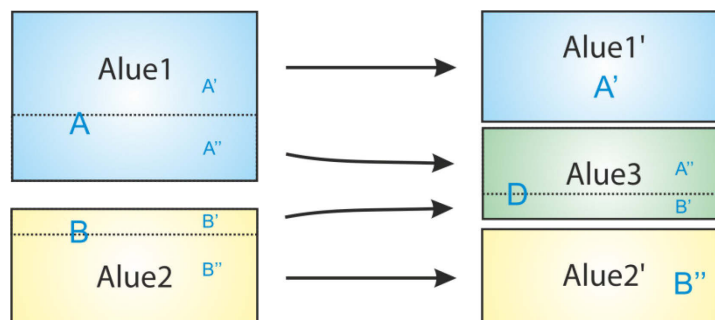
Aluesiirtoperustaminen

Uusi hallinnollinen alue voidaan perustaa kahdesta tai useammasta hallinnollisesta alueesta siirretyille osa-alueille. Tällöin tapahtuu kaksi erityyppistä muutosta yhdessä muutostapahtumassa: siirretään alueita jo olemassa olevista kunnista ja siirretyille alueille perustetaan uusi kunta. Tapahtuma eroaa pelkästä aluesiirrosta siten, että tuloksena on uusi hallinnollinen alue ja pelkästä perustamisesta siinä, että uuden kunnan alueet ovat peräisin jo olemassa olevista kunnista. Tällainen joukko yhdenaikaisia muutoksia on hankalampi muulla tavoin mallintaa, joten tapahtuma on otettu omaksi muutostyyppikseen.

Esimerkkinä aluesiirtoperustamisesta Imatran kauppala perustettiin vuonna 1948 kolmesta kunnasta siirretyille alueille. Yksi alueista oli Jääsken kunnasta alueluovutuksissa Suomen puolelle rajaa jäänyt palanen. Kuvassa 16 on kuntien rajat ennen ja jälkeen Imatran perustamisen. Kuva 17 esittää pinta-alojen ja sijaintitodennäköisyyksien kannalta katsottuna tilannetta, jossa kahdesta hallinnollisesta alueesta siirretyille alueille perustetaan uusi hallinnollinen alue.



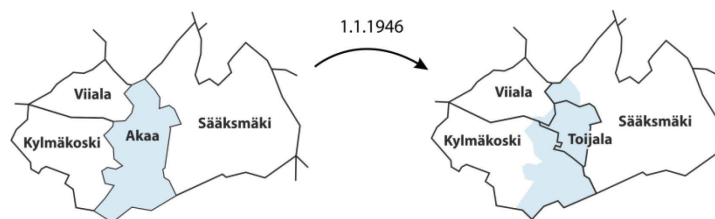
Kuva 16. Vuonna 1948 Imatran kunta perustettiin Ruokolahden, Joutsenon ja Jääsken kunnista siirretyille alueille. Jääsken kunta lakkasi samalla olemasta, koska suurin osa siitä oli jo vuoden 1940 alueluovutuksissa jäänyt Neuvostoliiton puolelle rajaa [23].



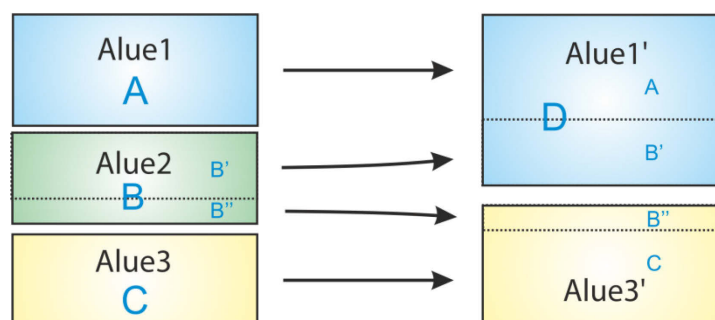
Kuva 17. Hallinnollisista alueista Alue1 ja Alue2 siirretään kummastakin osa uuteen perustettavaan Alue3:een. Alue1':n pinta-ala $A' = A - A''$, Alue2':n pinta-ala $B' = B - B''$ ja Alue3:n pinta-ala $C = A'' + B''$. Todennäköisyys, että Alue2:ssa sijainnut piste sijaitsee Alue3:ssa, on B''/B . Todennäköisyys, että Alue3:ssa sijaitseva piste on sijainnut Alue2:ssa, on B''/C .

Aluesiirtojakaantuminen (engl. division [6])

Aluesiirtojakaantuminen on käännteinen tapahtuma aluesiirtooperustamiselle. Hallinnollisen alueen maat siirretään usealle toiselle alueelle alkuperäisen hallinnollisen alueen lakatessa toimintansa. Kyseessä on siis muutostyyppien aluesiirto ja lakkauttaminen yhdistelmä. Jakaantumisesta tapahtuma eroaa siinä, että alueet siirtyvät ainakin osittain jo olemassa oleviin kuntiin. Kuvassa 18 Akaan kunta jaetaan osittain ympäröivien Viialan, Kylmäkosken ja Sääksmäen kesken. Entisen Akaan taajama-alueelle perustettiin samalla Toijalan kauppala. Kuva 19 esittää pinta-alojen jakaantumisen alueiden kesken.



Kuva 18. Akaan kunnan alue jaettiin vuoden 1946 alussa Toijalan, Kylmäkosken, Viialan ja Sääksmäen kesken. Kunnan keskustaajaman alueelle perustettiin Toijalan kauppal.



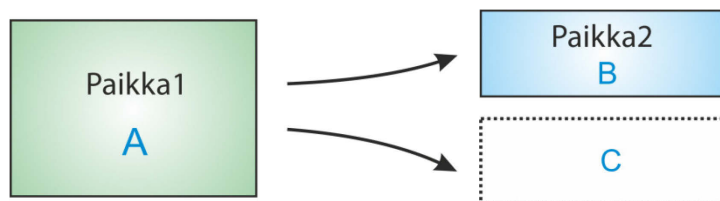
Kuva 19. Hallinnollinen alue Alue2 jakaantuu alueiden Alue1' ja Alue3' kesken. Todennäköisyys sille, että Alue1':ssa sijaitseva piste on kuulunut Alue2:een, on B'/D ja sille, että Alue2:ssa sijainnut piste on siirtynyt alueeseen Alue1' on B'/B .

Alueluovutus

Alueluovutus on valtioiden välisen rajan siirrosta johtuva hallinnollisen alueen luovuttaminen toisen valtion alueeksi (Kuva 20). Alueluovutus voi olla myös osittainen, jolloin vain osa hallinnollisesta alueesta siirtyy toiseen valtioon (Kuva 21). Suomi joutui vuonna 1940 luovuttamaan Moskovan rauhansopimuksen mukaisesti Neuvostoliitolle alueita Karjalasta ja Kainuun itäosista. Alueluovutukset koskivat yhteensä 71 kuntaa, joista luovutettiin 45 kuntaa kokonaan.



Kuva 20. Hallinnollinen alue Paikka1 luovutetaan kokonaan toisen valtion alueeksi ja lakkaa olemasta hallinnollinen alue lähtömaassaan.



Kuva 21. Osa hallinnollisesta alueesta Paikka1 luovutetaan toiselle valtiolle. Lähtömaahan jää alue Paikka2, jonka pinta-ala on luovutettua aluetta pienempi.

Aluesaaaminen

Aluesaaaminen on alueluovutukselle käänteinen tapahtuma, jossa toisesta valtiosta saadaan alue. Näin tapahtui Suomessa, kun Petsamon alue liitettiin Neuvosto-Venäjältä Suomeen Tarton rauhassa 1920. Alue liitettiin aluksi Inarin kuntaan, mutta pian perustettiin Petsamon kunta ja Petsamon lääni, joka sitten vuoden 1922 alusta liitettiin Oulun lääniin.

Kansainvälisesti katsottuna alueluovutus ja -saaaminen voitaisiin yhdistää yhdeksi muutostyypiksi, valtioiden väliseksi aluesiirroksi. Silloin ei otettaisi kantaa, mikä maa menettää ja mikä maa saa alueita. Kansallisessa tietojärjestelmässä tapausten tyypin erottaminen on kuitenkin mielekästä.



Kuva 22. Toisesta valtiosta saadaan alue, jolle perustetaan hallinnollinen alue Paikka1. Alueen pinta-ala on A.

Aluevuokraus ja aluevuokrauksen loppuminen

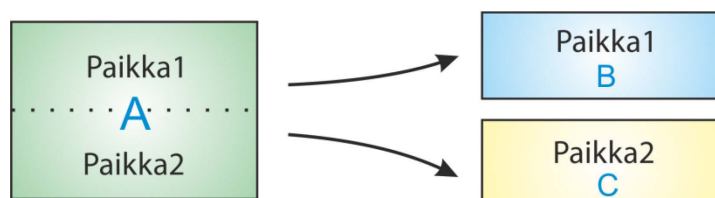
Aluevuokraus on valtioiden välinen jo alun perin määräaikaiseksi sovittu aluesiirto. Talvisodan rauhansopimuksessa 1940 Suomi joutui vuokraamaan Neuvostoliitolle Hangon ympäristöstä 30 vuodeksi 115 km² laajuisen alueen ja Moskovon välirauhassa 1944 Porkkalasta 380 km² maa-alueen 50 vuodeksi. Hangon alueen vuokraus päättyi käytännössä syksyllä 1944 ja virallisesti Pariisin rauhansopimuksen mukaisesti vuonna 1947. Porkkalan aluevuokraus lopetettiin puolestaan 1956.

Muodostuminen

Joidenkin vanhempien hallinnollisten alueiden perustamisesta ei ole jäänyt dokumentteja. Esimerkiksi keskiaikaisten linnaläänien perustamisajat ja -rajat eivät Porvoon linnalääniä lukuun ottamatta ole tiedossa, jolloin perustamisen sijasta käytetään nimitystä muodostuminen. Muutosten mallintamisen kannalta perustamisella ja muodostumisella ei ole muuta eroa kuin aika- ja rajatiedon epämääräisyys. Monesti on tiedossa vain aika-väli, jonka aikana muodostuminen on tapahtunut, mutta ei tarkkaa perustamisvuotta.

Rajanveto

Kahden historiallisen hallinnollisen alueen välisen rajan sijaintia ei ole välttämättä tarkkaan määritetty, jos se on esimerkiksi ollut asumaton erämaata tai vesistöä. Kun rajan kulku sitten sovitaan, sitä kutsutaan rajanvedoksi. Kuten kuvassa 23 esitetään, hallinnollisten alueiden pinta-alat on voitu tarkemmin määritellä vasta rajanvedon jälkeen. Tunnetuimpia rajanvetoja Suomen alueen historiassa on Ruotsin ja Novgorodin tasavallan välinen Pähkinäsaaren rauhan rajalinjaus vuodelta 1323.



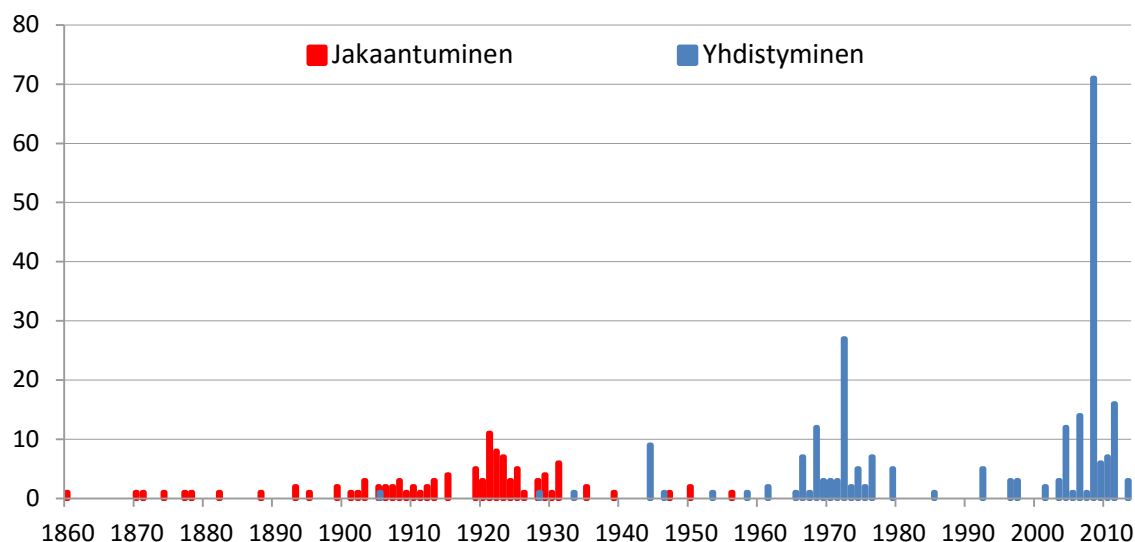
Kuva 23. Paikkojen Paikka1 ja Paikka2 rajan sijainti ei ole ollut tarkkaan määritelty. Rajanvedossa rajan sijainti määritellään ja samalla Paikka1:n ja Paikka2:n pinta-alat voidaan määritellä.

3.2 Muutosten vaikutus paikannimillä tapahtuvien hakujen tuloksiin

Muutosten ajallinen jakauma vaikuttaa merkittävästi siihen, miten paljon tapahtuneet muutokset aiheuttavat ongelmia esimerkiksi tietokantaan paikannimillä tehdyissä haussa. Muutosten lukumäärä ja tyyppi hakuhetken ja haettavan ajanhetken välillä vaikuttaa siihen, miten hyvän tuloksen haku tuottaa. Muutokset voivat myös kompensoida toisiaan. Esimerkiksi samalla alueella ensin tapahtunut jakaantuminen ja myöhemmin tapahtunut yhdistyminen voivat kompensoida toisensa. Esimerkiksi Juankosken kunta erotettiin Muuruvedestä vuonna 1926, kunnes vuonna 1971 kunnat liitettiin taas yhteen, mutta tällä kertaa nimellä Juankoski.

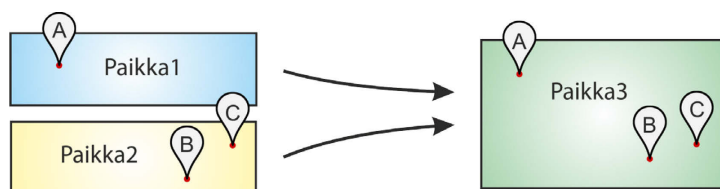
Kuva 24 esittää kunnissa tapahtuneiden jakaantumisten ja yhdistymisten vuosittaiset määrät. Kuntaliitoksia ei ole tapahtunut juuri lainkaan ennen 1940-lukua. 1940-luvusta

alkaen taas kuntien jakaantumiset ovat olleet hyvin harvinaisia. Jos kaaviota tulkitaan kunnan nimellä kuvailtujen aineistojen löydettävyyden kannalta, 1930-luvun jälkeiselle aineistolle on tärkeintä, että kuntaliitokset pystytään ottamaan huomioon hakutyökalun toiminnassa. Sen vanhemmalle aineistolle myös jakaantumisten huomioon ottamista tarvitaan, ellei tapahtunut jakaantuminen ole kompensoitunut myöhemmällä yhdistymisellä.

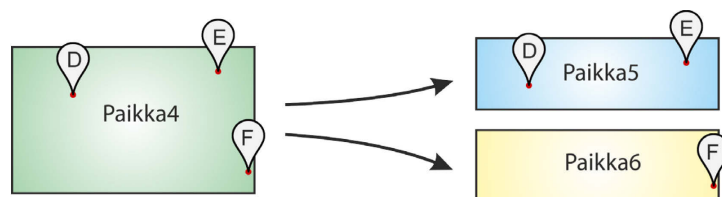


Kuva 24. Suomen kunnissa tapahtuneiden jakaantumisten ja yhdistymisten vuosittainen jakaantuminen.

Tutkitaan tyypillistä hakutilannetta, jossa haetaan vanhoja aineistoja nykyisellä hallinnollisella aluejaolla. Merkitsemme kaavakuviin kohteiden sijainteja kirjaimilla A-F. Kuva 25 esittää tilannetta, jossa kaksi aluetta, Paikka1 ja Paikka2 ovat liittyneet yhteen Paikka3:ksi. Tilanteessa on varmaa, että kaikki Paikka1:n ja Paikka2:n alueilla sijainneet kohteet (A–C) sijaitsevat nykyisen Paikka3:n alueella. Yhdistyminen on siksi helppo mallintaa aineistojen sijainnin kannalta. Kuvassa 26 Paikka4 jakaantuu Paikka5:ksi ja Paikka6:ksi. Tässä tilanteessa emme voi ilman tarkempia paikkatietoja tietää, löytyvätkö alun perin Paikka4:ssä sijainneet kohteet (D–F) nykyään Paikka5:stä vai Paikka6:sta. Tämä aiheuttaa haasteita hallinnollisten alueiden jakaantumisten mallinnukselle.

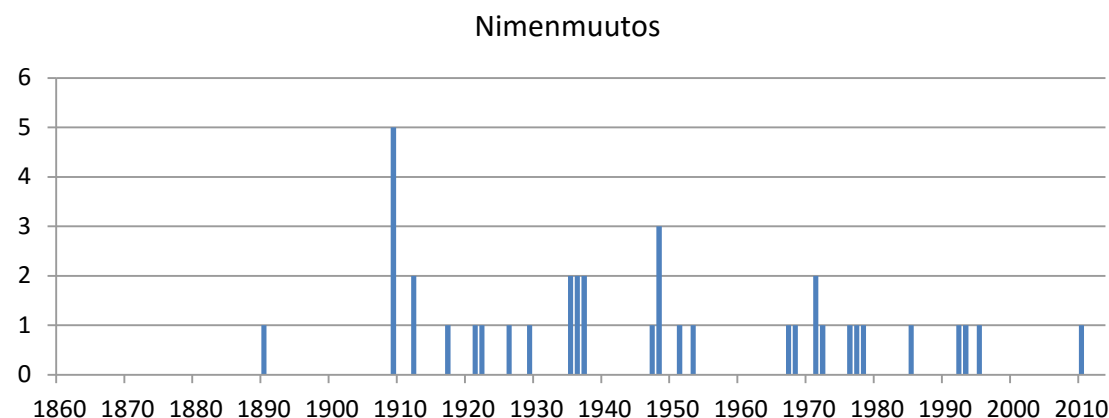


Kuva 25. Kaikki yhdistyvissä paikoissa Paikka1:ssä ja Paikka2:ssa olleet tutkimuspisteet (A, B ja C) sijoittuvat yhdistyneen Paikka3:n alueelle.

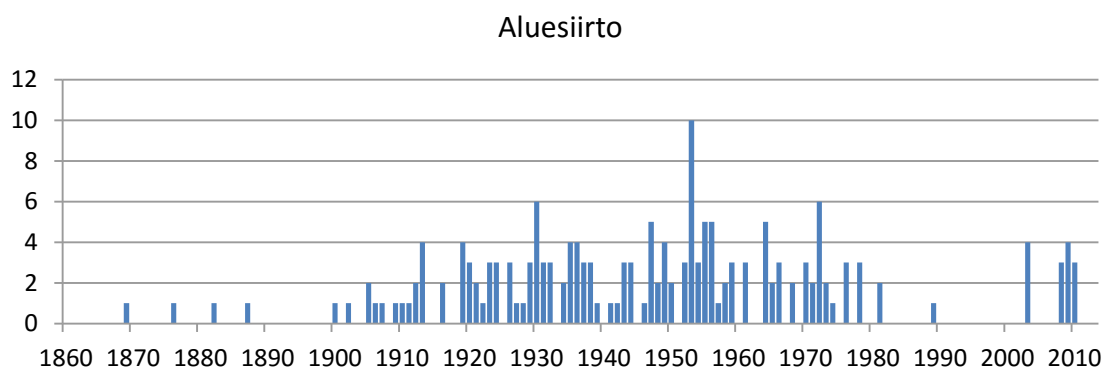


Kuva 26. Kun Paikka4 jakaantuu Paikka5:ksi ja Paikka6:ksi. Paikka4:ssä olleista kolmesta tutkimuspisteestä kaksi (D ja E) sijoittuu jakaantumisen jälkeen Paikka5:een ja yksi (F) Paikka6:een.

Kuntien nimenmuutokset jakaantuvat tasaisesti kautta historian lukuun ottamatta 1800-luvun loppupuolta, jolloin vasta oltiin sopeutumassa puhumaan kunnista pitäjien sijasta (Kuva 27). Nimenmuutosten sijaan ongelmia aiheuttavat tuon ajan kuntanimissä niiden vaihtelevat kirjoitusasut. Myös aluesiirtoja on tapahtunut kautta aikojen, mutta näyttää, että niitä on tapahtunut eniten 1950–1970-luvuilla kaupungistumisen kiihtyessä Suomessa (Kuva 28).



Kuva 27. Nimenmuutoksia on kunnissa tehty melko tasaisesti eri aikoina.



Kuva 28. Kuntien välisiä aluesiirtoja on tehty lähes vuosittain niin kauan kun kuntia on ollut.

Ei liene aiemmin tehty selvitystä siitä, haluavatko tietojärjestelmien käyttäjät hakea tietoja nykyisellä paikannimistöllä vai jollakin aiemmalla. Tekemässäni asiantuntijakyselyssä vastattiin esimerkiksi ”Asiakkaat käyttänevät hakusanana sitä kunnannimeä, jonka ovat oppineet tuntemaan.”. Koska saman sijainnin kunnannimi on saattanut vaih-

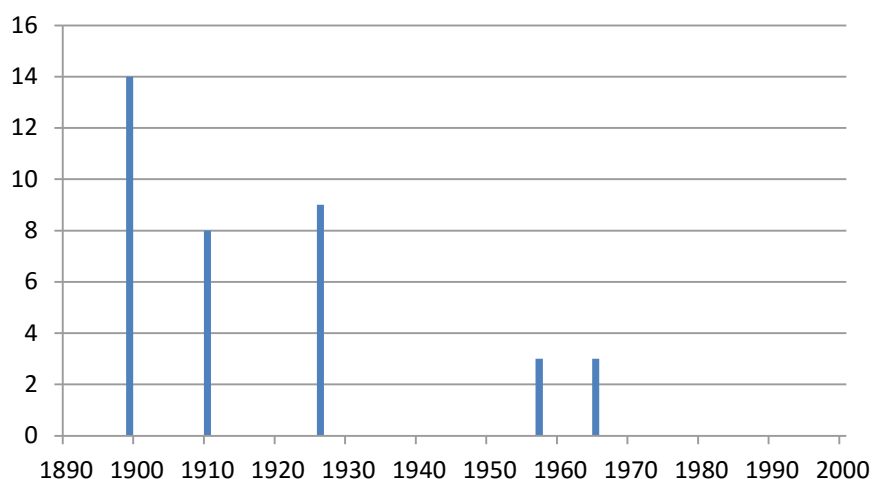
tua useaan kertaan historian kuluessa, vastaaja jatkaa: ”Olisi hyvä, jos paikkatiedot olisivat keskenään linkitetty myös ajallisesti toisiinsa, jolloin samalla haulilla voisi saada laajemmin osumia.” Olemassa oleva kuntajako lienee monille käyttäjille tuttu. Toisaalta kaikkein uusimmat kuntaliitokset eivät ole välttämättä vielä kaikkien tiedossa. Poikkeus ovat myös luovutetut alueet, joihin liittyvää aineistoa haetaan Suomessa hyvin todennäköisesti luovutushetken kuntajaon eikä nykyisen Venäläisen aluejaon perusteella. Toinen ymmärrettävä tapa hakea aineistoa on käyttää sen aikaista kuntajakoa, miltä ajalta haettava aineistokin on. Tämä vaatii tietenkin kuntamuutosten tuntemista tai selvittämistä.

3.3 Muita syitä muutoksiin alueiden pinta-aloissa

Hallinnollisten alueiden pinta-aloissa tapahtuneet muutokset kertovat, että alueiden rajoissa on tapahtunut muutoksia. Tämä pitää paikkansa, mikäli pinta-alojen mittaustiedot ovat luotettavia ja mikään muu syy ei vaikuta pinta-aloihin. Käytännössä alueiden pinta-alatiedoissa on virheellisyyttä, ja ainakin rannikkoalueella maankohoaminen aiheuttaa muutoksia pinta-aloihin, vaikkakin hitaasti.

Mittausvirheet

Arvioidakseni pinta-alamittausten tarkkuutta olen verrannut maapinta-alojen eriaikaisia mittaustietoja 80 sellaiselta kunnalta, joilla ei ole tietojemme mukaan tapahtunut aluemuutoksia vuosien 1899 ja 2000 välillä [55]. Mittaustiedot on poimittu viideltä eri vuodelta: 1899, 1910, 1926, 1957 ja 1965. Pinta-alatietojen lähteenä on käytetty Tilastollisen päätoimiston Suomen tilastollisia vuosikirjoja. Pinta-alatietoja on verrattu vuoden 2000 Maanmittauslaitoksen tietoihin, joiden on oletettu olevan nykyaikaisilla GIS-menetelmillä mitattuna tarkkoja ja luotettavia. Mitattujen pinta-alojen prosentuaalisten erojen keskiarvot nähdään pylväsdigrammissa kuvassa 29.

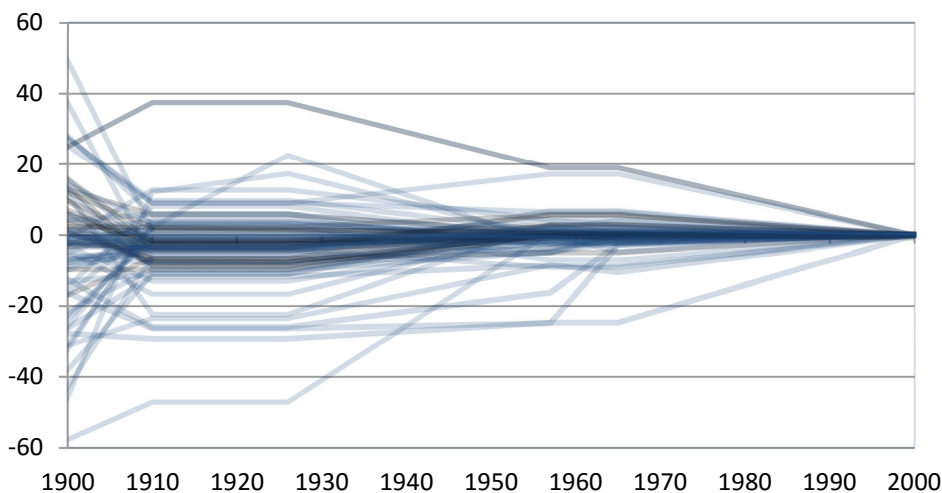


Kuva 29. Kahdeksankymmenen SAPO-ontologian mukaan rajoiltaan muuttumattomana pysyneen kunnan pinta-alamittausten keskimääräinen prosentuaalinen mittaero viitenä eri vuotena verrattuna vuoteen 2000.

Kuten on oletettavissa, kauemmaksi ajassa taaksepäin mentäessä mittausvirhe kasvaa. Keskimääräinen mittausvirhe on pysynyt alle kymmenessä prosentissa vuodesta 1910 alkaen. Kuvasta 30 nähdään, että mittaustulosten tarkkuuksissa on kuitenkin suuria eroja eri kuntien välillä. Joidenkin yksittäisten kuntien pinta-alojen mittausvirheet ovat hyvinkin suuria. Kuvasta voidaan myös havaita se, että pinta-alamittausten arvot pysyvät kahta poikkeusta lukuun ottamatta muuttumattomina vuosien 1910 ja 1925 välillä. Ilmeisesti pinta-aloja ei ole tuolla välillä tarkistettu muissa kunnissa, vaan on käytetty useana vuotena samoja mittaustuloksia, mikäli kuntien rajat ovat pysyneet muuttumattomina.

Koska muutos pinta-alatiedossa voi johtua myös todellisista muutoksista kunnan rajoissa, tietoa voidaan yleensä ottaen käyttää tapahtuneen muutoksen indikaattorina. Olen tarkastellut kahdenkymmenen eniten pinta-alaltaan muuttuneen kunnan rajatietoja karttapohjalla ja huomannut, että ainakin osassa näistä kunnista on kartalla havaittavissa rajasiirtoja, jotka puuttuvat vielä SAPO-ontologiasta. Virhe ei näissä tapauksissa ole pinta-alamittauksissa vaan siinä, ettei tietoa aluemuutoksista ole löytynyt mistään muutostietojen keräämisessä käytetyistä lähteistä. Tällaisen kohteet vaativat tarkempaa tutkimusta ja sen tuloksena mahdollista SAPOn täydentämistä. Osaan pinta-alamuutoksista ei ole löytynyt mitään selitystä ja niitä täytyy käsitellä vain mittausvirheinä.

Jos pinta-alamuutoksia käytetään aluemuutosten indikaattoreina, osa pienistä aluemuutoksista jää huomaamatta, koska pinta-alamittausvirheiden suuruus kätkee ne alleen. Alle kymmenen prosentin muutoksia pinta-alamittauksissa on paljon. Siksi niiden kaikkien tarkastaminen mahdollisina aluesiirtoina on käytännössä hyvin työlästä.



Kuva 30. Kahdeksankymmenen SAPO-ontologian mukaan rajoiltaan muuttumattomana pysyneen kunnan prosentuaalisten mittausvirheiden jakautuminen.

Maankohoaminen

Jääkaudella Suomen rannikkoalueiden maankamara painui jäätikön painosta alaspäin. Jääkauden jälkeen maa on noussut koko Suomen alueella. Rannikkoalueilla maanousu aiheuttaa rannan siirtymistä kauemmas merelle ja sitä kautta maapinta-alan kasvamista. Mittausten mukaan nousu on tällä hetkellä maksimissaan 9 millimetriä vuodessa [19]. Maanousu on nopeinta Pohjanmaan rannikolla ja hitaampaa Suomenlahdella. Maankohoamisen vaikutus alueen pinta-alaan riippuu kohoamisnopeudesta ja rannan jyrkkyy-

destä. Mitä loivempi ranta on, sitä nopeammin sitä paljastuu veden alta maan noustessa. Loivassa rannassa myös sedimentoituminen nopeuttaa maankohoamista osaltaan.

Maankohoaminen voi periaatteessa aiheuttaa virhettä pinta-alan perusteella tehtäviin sijainnin todennäköisyysarvoihin. Tämä johtuu siitä, että vertailulukuina laskuissa käytetään maapinta-aloja. Käytännössä vaikutus on kuitenkin pieni. Esimerkiksi Pohjanmaan rannikolla sijainneen Maksamaan kunnan pinta-alan lasketaan kasvaneen yhdeksän hehtaarin vuosivauhtia [34]. Maksamaan maapinta-ala oli 152 km^2 ennen liittymistä Vöyrin kuntaan, joten vuotuinen pinta-alan kasvu oli prosentuaalisesti pieni (0,6 promillea). Pidemmällä aikavälillä katsottuna maankohoaminen on kyllä vaikuttanut rannikkoalueen kuntiin. Kokonaisia kaupunkeja, esimerkiksi Pori ja Vaasa, on jouduttu historian kuluessa siirtämään, koska rannansiirtyminen on vienyt niiltä vesiyhteyden merelle.

4 Muutostietojen mallintaminen ontologiassa

Suomen ajallista paikkaontologiaa (SAPO) alettiin kehittää osana Suomen semanttisen webin ontologiat (FinnONTO) hanketta vuonna 2005. Kauppinen ja Hyvösen julkaisuissa [22] kuvataan ontologian tuottamisessa käytetyt menetelmät, määritellään ontologian käytöstä saatavaa hyötyä aineistoja hakiessa sekä esitellään ontologiaa käyttävä semanttinen verkkopalvelu Kulttuurisampo.

Ontologian lähtöaineistona käytetään selvittämiäni Suomen hallinnollisissa alueissa tapahtuneita muutostietoja, jotka kuvattiin edellisessä kappaleessa. Muutostiedot on koottu Suomen alueella sijainneista kunnista, lääneistä ja valtioista. SAPO-ontologiassa keskityttiin ennen kaikkea ajallisten kuntien mallintamiseen, koska ne nähtiin aineistojen kuvailun kannalta tärkeimmiksi hallinnollisiksi alueiksi. Ne soveltuivat myös parhaiten muutosten mallinnusten testaukseen, koska niissä on tapahtunut hyvin monenlaisia muutoksia. Myös ajallisten läänien ja niiden hierarkkisten peittävyysuhteiden mallintamista suhteessa kuntiin kokeiltiin.

SAPO käyttää Resource Description Frameworkiin (RDF) ja RDF-skeemaan (RDFS) perustuvaa Simple Knowledge Organization System (SKOS) –tietomallia [49] ja Linked Data -tekniikoita tietomallin toteutuksessa [56]. Nämä ovat W3C:n suosittelemia teknologioita ja käytössä myös muissa FinnONTO-hankkeen tuloksena julkaistuissa ontologioissa. SKOS on erityisesti termistöjen julkaisemista varten kehitetty RDF-pohjainen formaatti.

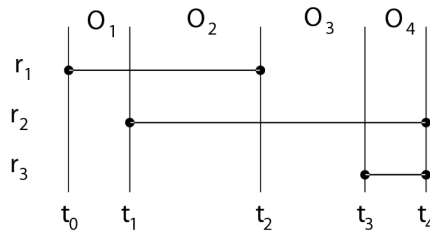
4.1 Muutoksia kuvaava ontologiasarja

Ajallisen paikkaontologian (engl. spatio-temporal ontology) kehittämisen tavoitteena on mahdollistaa kohteiden tarkka annotointi, semanttinen haku, semanttinen linkitys ja koneellisen päättelyn avulla tapahtuva rikastaminen [10]. Ontologista rakennetta (esimerkiksi osa/kokonaisuus -suhteita) voidaan käyttää apuna myös siinä, kuinka aineisto esitetään käyttöliittymässä.

Ajallisen paikkaontologian tuottaminen vaatii paikkakäsitteiden spatiaalisten ominaisuuksien lisäksi huomioimaan neljännen ulottuvuuden, ajan. Tarkoituksena on luoda ikään kuin sarja useita toisiaan seuraavia ontologioita, jossa peräkkäiset ontologiat poikkeavat toisistaan vain kulloinkin muuttuneiden paikkatietojen osalta [22]. Kukin ontologia kuvaa aina yhtä muuttumattomana pysyvää ajanjaksoa. Ontologia luodaan kartoittamalla ajanhetket, jolloin alueiden rajoissa tai nimissä on tapahtunut muutoksia. Jos muutoksia on tapahtunut, mallinnetaan niiden vaikutukset alueiden keskinäisiin ominaisuuksiin ja otetaan käyttöön uusi muuttunut tilannetta kuvaava ontologian versio.

Kuva 31 [22] esittää neljää toisiaan seuraavaa ontologiaa O_1 – O_4 . Muutoshetkellä t_0 – t_4 tapahtuva muutos yhdessä tai useammassa ontologioiden resurssissa r_1 – r_3 aiheuttaa uuden ontologian luomisen. Jokainen resurssi kuuluu vähintään yhteen ontologiasarjan ontologioista ja on olemassa vähintään seuraavaan muutoshetkeen ja ontologiaversioon saakka.

Kukin yksittäinen ontologia koostuu kaikista niistä resursseista, jotka ovat olemassa sen käsittämän ajanjakson aikana. Aineistojen kuvailussa käytetään tilanteeseen sopivaa ontologian ajallista versiota. Vastaavasti aineistoja hakiessa voidaan käyttää toista ontologian ajallista versiota, jolloin ontologiasarja kytkee käytetyt resurssit halutun aikajakson yli toisiinsa [22].



Kuva 31. Muuttuvia resursseja r_k kuvaava ontologiasarja O_i . Muutokset tapahtuvat muutoshetkinä t_j [22].

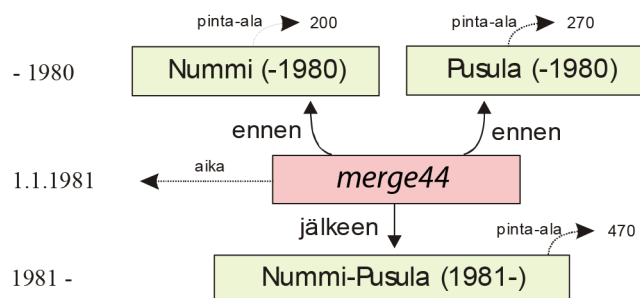
Näiden ajallisesti määriteltyjen tilannekuvakäsitteiden lisäksi tarvitaan käsitteitä, jotka kuvaavat samoja paikkakäsitteitä myös koko niiden olemassaolon ajalta, esimerkiksi Tampere kaikkina aikoina. Tätä tarvetta varten Kauppinen & Hyvönen esittelevät käsitteen aikamato (engl. space worm) [22]. Tällainen aikamato koostuu kaikista Tampereen ajallisista kunnista Tampereen kaupungin perustamisesta tähän päivään ja tulevaisuuteen saakka – niin kauan kuin Tampereen kaupunki on olemassa.

4.2 Spatiaaliset muutossillat

Kauppinen ja Hyvönen esittelevät julkaisussaan [22] käsitteen *muutossilta*, joka kuvaa muutostapahtuman vaikutuksia muutoksessa mukana olevien alueiden spatiaalisiin suhteisiin. Muutossilta kuvaa esimerkiksi, mitä alueita muutoksen takia lakkaa olemaan olemassa, mitä uusia alueita muutoksessa syntyy ja kuinka paljon ennen muutosta ja sen jälkeen olevat alueet peittävät toisiaan.

Jokaiselle perusmuutostyypille määritellään oma RDFS [2] muutossiltaluokkansa. Jokainen yksittäinen muutos on instanssi tästä luokasta. Muutossiltaluokkia Kauppinen ja Hyvönen kertovat [22] tarvittavan ainakin uuden alueen eli resurssin luomiselle (*addition*), nimenmuutokselle (*usedtobe*), lakkauttamiselle (*removal*), alueiden yhdistymiselle (*merged*) ja jakaantumiselle (*split*).

Kuvassa 32 on esimerkki muutossillasta, joka on tyyppiä *merged* eli se kuvaa alueiden yhdistämistä. 1.1.1981 Nummen ja Pusulan kunnat yhdistyivät Nummi-Pusulan kunnaksi [54]. Muutossillan ominaisuudet *ennen* kuvaavat tilannetta ennen yhdistymistä ja *jälkeen* muutoksen jälkeistä tilannetta. Ominaisuus *aika* kertoo muutoksen tapahtuma-ajan. Resurssien pinta-alojen summien ennen muutosta ja muutoksen jälkeen täytyy olla aina samat muutossilloissa, jotka ovat tyyppiä *usedtobe*, *merged* tai *split*.



Kuva 32. Nummen ja Pusulan kuntien yhdistämistä kuvaava RDFS-muutossilta [54].

Kuvan 32 muutossilta voidaan esittää RDF-kolmikoiden [2] avulla:

Subjekti	Predikaatti	Objekti
change: merge44	change:before	sapo:Nummi (-1980)
change: merge44	change:before	sapo:Pusula (-1980)
change: merge44	change:after	sapo:Nummi-Pusula (1981-)
change: merge44	dcterms:date	"1981-01-01"

Tässä change viittaa muutostyyppien nimiavaruuteen Change Vocabulary⁹, dcterms viittaa Dublin Coren termistöön¹⁰ ja sapo SAPOn paikkakäsitteisiin.

4.3 Sijainnin todennäköisyyden laskeva päättelykone

Kauppisen ja Hyvösen menetelmän avulla eriaikaisten alueiden päällekkäisyystietoja voidaan laskea ja käyttää hyväksi alueilla sijaitsevien kohteiden sijaintien todennäköisyyksien laskemisessa. Voidaan esimerkiksi määrittää, miten suurella todennäköisyydellä suurimmalta osaltaan Neuvostoliitolle luovutetulla alueella sijainneessa Jääsken kunnassa valmistetun paimentorven valmistuspaikka voisi sijaita nykyisen Imatran kaupungin alueella.

Todennäköisyyksien suuruudet voidaan laskea mallintamalla ensin muutossillat, laskemalla niiden avulla paikalliset peittävyudet ja lopulta laskemalla paikallisten peittävyuksien avulla globaalit peittävyudet [22]. Peittävyysarvot ovat desimaalilukuja väliltä nollasta yhteen. Arvolla nolla peittävyyttä ei ole lainkaan ja arvolla yksi alue peittää toista aluetta täydellisesti. Luvut lasketaan kahden eriaikaisen paikan kesken ajallisesti molempiin suuntiin: *peittää* (engl. covers) kuvaa kuinka paljon uudempi alue peittää vanhempaa aluetta ja *tulee peitettyksi* (is covered by) paljonko vanhempi alue peittää uudempaa. Esimerkiksi nykyinen Imatra peittää entistä Jääsken kuntaa arvolla 0.095 ja tulee peitettyksi arvolla 0.258. Peittävyystietoa voidaan suoraan käyttää sijainnin todennäköisyysarvona sillä oletuksella, että haettava aineisto on alueelle tasaisesti jakaantunut. Jostain tarkemmin määrittelemättömästä paikasta Jääsken alueelta löytyneet paimentorven löytöpaikan todennäköisyys sijaita nykyisessä Imatrassa on siis 0.095 eli 9,5 prosenttia.

Joissain tilanteissa todennäköisyyksien laskeminen voi olla mutkikasta, koska varsinkin pidemmillä aikaväleillä peittävyksiä voi tulla useita muutosketjuja pitkin. Eri-tyisen hankala on tilanne, jos jakaantuminen seuraa yhdistymistä [22]. Ilman lisämäärittelyä ei voida päätellä, miten uusi jakaantuminen tapahtuu suhteessa aiempiin yhdistyneisiin alueisiin. Jos raja kulkee aiemmassa paikassa, ei ole mitään ongelmaa. Jos raja kulkee täysin uudessa paikassa, tieto alueiden jakaantumisesta on annettava manuaalisesti.

Kuten kuvasta 24 nähtiin, Suomessa jakaantumisia on tapahtunut pääasiassa ennen 1940-lukua ja yhdistymisiä vasta sen jälkeen. Koko Suomen kuntamuutoshistoriassa on vain yksi tapaus, jossa yhdistymistä on samalla alueella seurannut jakaantuminen. Vuonna 1929 Mustio yhdistettiin Karjaan maalaiskuntaan ja heti seuraavana vuonna 1930 Karjaan kauppala erosi maalaiskunnasta. Tämäkin ongelmallinen muutosketju

⁹ <http://linkedearth.org/change/ns/>

¹⁰ <http://purl.org/dc/terms/#>

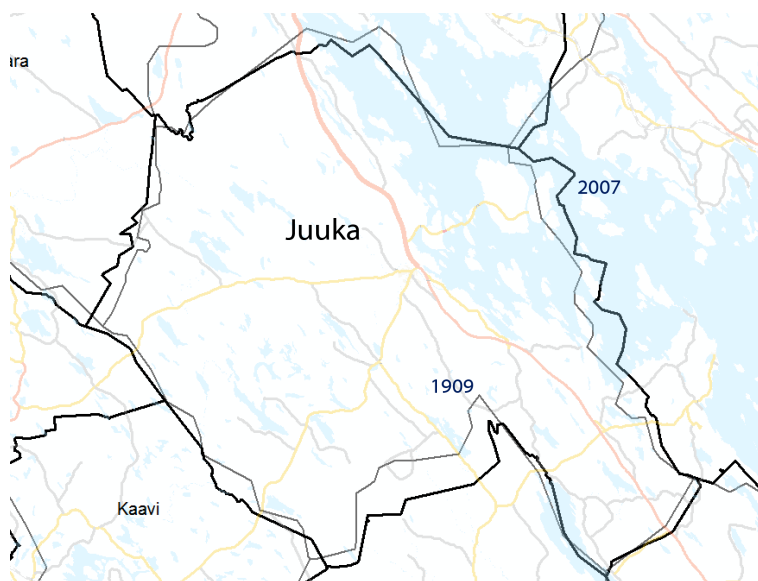
ikään kuin korjaantui vuonna 1969, kun Karjaan maalaiskunta ja Karjaan kauppala yhdistyivät ja peittävyystietojen laskeminen ketjun yli automaattisesti tuli taas mahdolliseksi.

Koska menetelmä käyttää laskennassa alueiden pinta-alasuhteita, se ei ota huomioon muita sijaintitodennäköisyyksiin vaikuttavia tekijöitä kuten asutuksen jakaantumista tai kulkuyhteyksiä. Parhaiten menetelmä toimii siis aineistolle, joka on jakaantunut tasaisesti kohdealueelle.

4.4 Hallinnollisten alueiden rajapolygonit

Hallinnollisten alueiden rajoilla on ollut suuri merkitys ihmisten elämään. Rajat onkin merkitty kartoille, kun uusia kartoituksia on tehty tai olemassa olevia kartoja päivitetty. Rajapolygonien avulla on mahdollista rekonstruoida hallinnollisissa alueissa tapahtuneita ajallisia muutoksia kappaleessa 2.4 kuvatuilla menetelmillä. Suomessa kartoitusten ajallinen tiheys ei riitä kuvaamaan kaikkia ajallisia kuntia. Olemassa olevien karttojen rajatiedot yhdistämällä muualta saatuihin muutostietoihin on kuitenkin mahdollista tuottaa varsin kattava rajamuutossarja kunnista ainakin 1900-luvun alusta nykyhetkeen.

SAPOn kehitystyön yhteydessä digitoitiin Geologian tutkimuskeskuksen työnä Suomen kuntien rajapolygonit tilannekuvina vuosilta 1909 ja 1960. Vuoden 2007 rajat saatiin Maanmittauslaitoksesta suoraan digitaalisessa muodossa. Lisäksi digitoitiin Kaakkois-Suomi vuodelta 1940 ja Petsamo vuodelta 1927. Vuosien 1909 ja 1960 digitoinnit tehtiin suurimittakaavaisista Suomen kartaston koko Suomen kattavista kartoista. Kuvasta 33 nähdään ongelma, joka syntyy, kun eriaikaisia ja eri lähteistä olevia rajatietoja digitoidaan. Digitointien perusteella näyttäisi siltä, että rajoissa on tapahtunut runsaasti muutoksia vuosien 1909 ja 2007 välillä. Todellisuudessa erot polygonien sijainneissa johtuvat vanhemmassa kartassa olevista virheistä ja digitoitujen karttojen huomattavasta mitta-kaavaeroista. Toinen ongelma tämänkaltaisessa tilannekuvamenetelmässä on se, että samat muuttumattominakin pysyneet rajat digitoidaan joka tilannekuvassa turhaan aina uudestaan, mikä vaatii turhaa työtä ja tuottaa päällekkäistä tietoa [35].



Kuva 33. Vuosien 1909 ja 2007 välillä todellisuudessa muuttumattomina pysyneissä rajoissa näyttäisi eriaikaisten digitoitien perusteella tapahtuneen muutoksia.

SAPOn julkaisuversiossa ONKI-ontologiapalvelussa¹¹ näytetään käytettävissä olevat rajapolygonit Google Maps -karttapohjalla. Eriaikaisista digitoiduista rajapolygoneista valitaan näytettäväksi uusin ja tarkin versio, jos samasta ajallisesta kunnasta on useita digitointeja. Lisäksi rajapolygoneja käytetään Kulttuurisampo-palvelussa¹² apuna erilaisten historiallisten kohteiden sijaintien esittämisessä [25]. Valitsemalla jokin historiallinen kunta, palvelu näyttää sen kattamalta alueelta eri aikoina kerättyä kulttuuri-historiallista esineistöä. Kulttuurisammossa on myös mahdollisuus tuoda samalle karttapohjalle yhtä aikaa useita saman kunnan eriaikaisia rajoja. Siten voidaan verrata helposti, miten aiemmat rajat ovat eronneet nykyisistä.

4.5 SAPOn ylläpito ja julkaiseminen

Kaikkia SAPOn muutostietoja ylläpidetään Excel-taulukossa (Kuva 34). Tärkeimmät taulukkoon kerätyt tiedot ovat ajallisen paikan nimi, muutoksen ajankohta ja muutosmuoto sekä vapaamuotoinen kuvaus tapahtuneesta muutoksesta. Lisäksi taulukkoon kootaan kuntien pinta-alatiedot eri ajoilta, nimet suomeksi ja ruotsiksi sekä eri aikoina ja eri lähteissä käytettyjä vaihtoehtoisia nimiä. Liitteessä 1 luetellaan muutostietojen selvittämisessä käytettyjä lähteitä.

1	Kunta	Vuosi	Muutosmuoto	Huomautukset	Aikupvm
2	Ahlainen	1908	Jakaantuminen	Ahlainen erotettiin Ulvilasta 1908	1908-01-01
3	Ahlainen	1972	Yhdistyminen	Ahlainen liitettiin Porin 1972	1972-01-01
4	Aitolahti	1923	Jakaantuminen	Aitolahti erotettiin Messukylästä 5.7.1923	1923-01-01
5	Aitolahti	1966	Yhdistyminen	Aitolahti liitettiin Tampereeseen 1966	1966-01-01
6	Akaa	1870	Perustaminen	Akaa perustettiin 1870	1870-01-01
7	Akaa	1895	Jakaantuminen	Kylmäkoski erotettiin Akaasta 1895	1895-01-01
8	Akaa	1932	Alueiirto	Akaasta siirrettiin Viialaa perustettaessa osia 1932	1932-01-01
9	Akaa	1946	Alueiirtojakaantuminen	Akaa jaettiin 1946 (Kylmäkoski, Sääksmäki, Toijala, Viiala)	1946-01-01
10	Akaa_2	2007	Yhdistyminen	Toijala ja Viiala yhdistettiin Akaaksi 2007	2007-01-01
11	Akaa_2	2011	Yhdistyminen	Kylmäkoski liitettiin Akaaseen 2011	2011-01-01
12	Alahärmä	1867	Perustaminen	Alahärmä perustettiin 1867	1867-01-01
13	Alahärmä	2009	Yhdistyminen	Alahärmä, Korttesjärvi ja Ylihärmä liitettiin Kauhavaan 2009	2009-01-01
14	Alajärvi	1868	Perustaminen	Alajärvi perustettiin 1868	1868-01-01
15	Alajärvi	2009	Yhdistyminen	Lehtimäki liitettiin Alajärveen 2009	2009-01-01
16	Alastaro	1869	Perustaminen	Alastaro perustettiin 1869	1869-01-01
17	Alastaro	2009	Yhdistyminen	Alastaro ja Mellilä liitettiin Loimaaseen 2009	2009-01-01
18	Alatornio	1865	Perustaminen	Perustamisvuosi ei tiedossa. Käytetään oletuksena vuotta 1865.	1865-02-06
19	Alatornio	1957	Alueiirto	Alatorniosta siirrettiin Tornioon Putaa ja Röyttä 1957	1957-01-01
20	Alatornio	1973	Yhdistyminen	Alatornio liitettiin Tornioon 1973	1973-01-01
21	Alaveteli	1896	Perustaminen	Alaveteli perustettiin 1896	1896-01-01
22	Alaveteli	1969	Yhdistyminen	Alaveteli liitettiin Kruunupyyn 1969	1969-01-01
23	Alavieska	1879	Perustaminen	Alavieska perustettiin 1879	1879-01-01
24	Alavus	1865	Perustaminen	Alavus perustettiin 1865	1865-02-06
25	Alavus	1973	Alueiirto	Alavudelta siirrettiin Nurmoon alueita 1973	1973-01-01
26	Alavus	2013	Yhdistyminen	Töysä liitettiin Alavukseen 2013	2013-01-01
27	Angeliemi	1916	Jakaantuminen	Angeliemi erotettiin Halikosta 1916.	1916-01-01

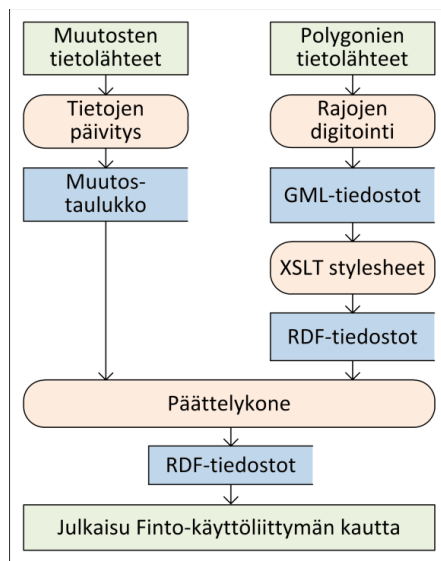
Kuva 34. Ote taulukosta, jossa ylläpidetään SAPOn lähdetietoja.

Kuva 35 näyttää SAPOn ylläpitoprosessin vaiheet. Taulukosta tiedot luetaan Tomi Kauppinen kehittämälle Java-ohjelmointikielellä ja Jena Semantic Web Frameworkin [3] avulla toteutetulle päättelykoneelle. Päättelykone tuottaa muutostietojen avulla SKOS-muotoa olevat ontologiatiedostot.

Päättelykone käyttää toisena lähdeaineistona GML-formaatissa tallennettuja rajapolygoneja. Ne muunnetaan XSLT-tiedoston [26] (Extensible Stylesheet Language Transformations) avulla RDF-muotoon. Päättelykone hakee kullekin ajalliselle kunnalle uusimman käytettävissä olevan rajapolygonin. Esimerkiksi kuvan 33 polygoneista päättelykone valitsisi Juuan kunnalle vuoden 2007 polygonin vuoden 1909 sijasta, koska sillä aikavälillä Juuan kunnassa ei ole tapahtunut aluemuutoksia.

¹¹ <https://onki.fi/>

¹² <http://www.kulttuurisampo.fi/>



Kuva 36. Kaaviokuva SAPOn ylläpitoprosesseista [56].

Alla on päätelykoneen tuottama RDF Turtle sarjallistus Kruunupyyn 1.1.1896 alka-
neesta ajallisesta kunnasta. Turtle on sekä kone- että ihmisluettava formaatti, jonka
avulla pystytään tiiviisti kuvaamaan RDF-kolmikoita (subjekti, predikaatti ja objekti)
sanallisesti.

```

@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xs:       <http://www.w3.org/2001/XMLSchema#> .
@prefix skos:     <http://www.w3.org/2004/02/skos/core#> .
@prefix sapometa: <http://linkedearth.org/meta/sapo#> .
@prefix change:   <http://linkedearth.org/change/ns/> .
sapo:Kruunupyy_1896_01_01
  rdf:type sapometa:Kunta ;
  skos:altLabel "Kruununkylä"@fi ;
  skos:prefLabel "Kruunupyy (1896-01-01)"@fi , "Kronoby (1896-01-01)"@sv ;
  change:begindate "1896-01-01"^^xs:date ;
  sapometa:overlaps sapo:Kruunupyy_1932_01_01 , sapo:Oja_1932_01_01 ;
  sapometa:overlaps sapo:Kruunupyy_1865_02_06 ;
  change:enddate "1931-12-31"^^xs:date ;
  sapometa:hasUnionOf sapo:Kruunupyy ;
  sapometa:sizeofLandArea "317.0"^^xs:float .
  
```

Kuvaus sisältää seuraavia tietoja:

Ensimmäisillä riveillä luetellaan käytettyjen nimiavaruuksien etuliitteet (engl. prefix).
Etuliitteiden ansiosta käytettyjen resurssien URI-tunnuksia ei tarvitse kirjoittaa moneen
kertaan, mikä lyhentää ja selventää kuvausta.

sapo:Kruunupyy_1896_01_01

Tietue koskee SAPO-ontologian käsitettä Kruunupyy_1896_01_01.

rdf:type sapometa:Kunta ;

Kyseessä on käsite, joka on tyyppiä Kunta

skos:altLabel "Kruununkylä"@fi ;

Kuntaan on voitu viitata myös rinnakkaisnimellä Kruununkylä.

skos:prefLabel "Kruunupyy (1896-01-01)"@fi , "Kronoby (1896-01-01)"@sv ;

Kunnan suositellut nimet ovat suomeksi Kruunupyy ja ruotsiksi Kronoby.

```
change:begindate "1896-01-01"^^xs:date ;
```

Tällä viitataan change-sanastosta löytyvään käsitteeseen begindate, jolla tarkoitetaan ajallisen kohteen alkupäivämäärää.

```
sapometa:overlaps sapo:Kruunupyy_1932_01_01 , sapo:Oja_1932_01_01 ;
sapometa:overlaps sapo:Kruunupyy_1865_02_06 ;
```

Minkä muiden ajallisten kuntien kanssa Kruunupyy_1896_01_01:llä on päällekkäisiä alueita.

```
change:enddate "1931-12-31"^^xs:date ;
```

Ajallisen kunnan viimeinen olemassaolopäivä.

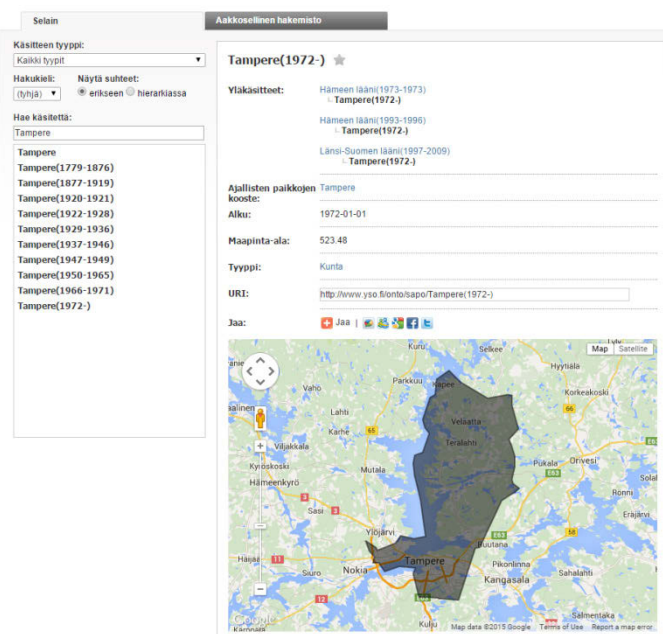
```
sapometa:hasUnionOf sapo:Kruunupyy ;
```

Pelkkä kunnan nimi ilman aikamäärettä viittaa ajattomaan Kruunupyyhyn, joka kattaa aikamaton kaikki ajalliset Kruunupyyt.

```
sapometa:sizeofLandArea "317.0"^^xs:float .
```

Ajallisen kunnan Kruunupyy_1896_01_01 pinta-ala neliökilometriä.

SAPO on julkaistu Semanttisen laskennan tutkimusryhmän¹³ (SeCo) ylläpitämässä ONKI-ontologiapalvelussa. Kuvassa 37 palvelusta on haettu kuntaa Tampere. Vasemmassa sarakkeessa nähdään Tampereen kaikki ajalliset kunnat ja koostekäsité Tampere ilman aikamäärettä. Oikeassa sarakkeessa nähdään, mihin ajallisiin lääneihin ajallinen kunta Tampere(1972-) on kuulunut, muuta tietoa ajallisesta Tampereesta ja alimman sen rajat Googlen karttapohjalla.



Kuva 37. Tampereen ajalliset kunnat ONKI-palvelussa. Valittuna on Tampere(1972-), jonka rajoista on olemassa myös karttapolygoni.

¹³ <http://www.seco.tkk.fi/>

Vuoden 2016 aikana muutamilta osin täydennetty ja päivitetty SAPO julkaistaan Kansalliskirjaston Finto-palvelussa. Uusina ominaisuuksina palvelu sisältää paikkojen eriaikaisia rinnakkaisnimiä ja myös ruotsinkieliset nimet. Suurimpana muutoksena uudessa versiossa käsitteiden URI-tunnisteet on muutettu esimerkiksi muodosta [http://www.yso.fi/onto/sapo/Tampere\(1972-\)](http://www.yso.fi/onto/sapo/Tampere(1972-)) muotoon http://www.yso.fi/onto/sapo/Tampere_1972_01_01. Muutoksen tarkoituksena on parantaa URI:n pysyvyyttä. Aiemmin joka kerta uuden muutoksen tapahtuessa, muutoksessa mukana olleiden ajallisten paikkojen URI:t jouduttiin muuttamaan lisäämällä URI:n loppuvuosi. Uudenmuotoisissa URI-tunnisteissa on pelkästään kunkin ajallisen paikan alkupäiväys, jolloin uusi muutos ei vaikuta URI:iin. Päiväyksen käyttö URI:ssa tekee mahdolliseksi myös useiden samana vuonna tapahtuneiden muutosten kuvaamisen omilla ajallisilla paikoillaan. Tähän ei tosin toistaiseksi ole ollut Suomen hallinnollisissa alueissa tarvetta.

Kuvassa 38 on esimerkki, miltä SAPO näyttää Finto-palvelussa. Aakkosellisesta listasta on kuvassa valittu ajallinen kunta Iisalmen mlk (1922-01-01). Käsitteen voi palvelusta ladata omalle koneelle RDF/XML, Turtle tai JSON-LD -formaateissa. Paikkojen rajapolygoneja Finto-palvelun ensimmäinen versio ei tule vielä sisältämään, koska tarvittavaa karttakäyttöliittymää ei palvelussa ole vielä käytettävissä. Myös paikkojen peittävyysuhteet tulevat puuttumaan tässä vaiheessa.

The screenshot shows the Finto SAPO - Suomen ajallinen paikkaontologia interface. The top navigation bar includes the Finto logo and links for Sanastot, Tietoja, Palaute, and Ohje. The main header displays 'SAPO - Suomen ajallinen paikkaontologia' and a language selector set to 'suomi'. The left sidebar shows a hierarchical list of terms, with 'Iisalmen mlk (1922-01-01)' selected. The main content area displays the details for 'Iisalmen mlk (1922-01-01)', including its type (Kunta), existence dates (1.1.1922 to 31.12.1969), and URI (http://www.yso.fi/onto/sapo/iisalmen_mlk_1922_01_01). The interface also provides options to download the concept in RDF/XML or Turtle format.

Kuva 38. Iisalmen maalaiskunta Finton ontologiapalvelun kehitysversiossa kesällä 2015

Koska hallinnollisissa alueissa tapahtuu muutoksia lähes vuosittain, SAPOa täytyy täydentää vähintään kerran vuodessa. Sopiva hetki täydennykselle on vuosittain marras-joulukuussa, jolloin kaikki seuraavassa vuodenvaihteessa tapahtuvat muutokset ovat tiedossa. Kun SAPO vuonna 2015 julkaistaan Finto-palvelussa, on todennäköistä, että Kansalliskirjasto huolehtii jatkossa myös siitä, että ontologia pysyy ajan tasalla.

5 Muutostietojen mallintaminen valokuvatietokannassa

Kiinnostukseni paikkatiedon ajallisiin muutoksiin alkoi täyttyessäni metatietoja Geologian tutkimuskeskuksen vanhojen valokuvien tietokantaan. Kuvateksteissä oli runsaasti kuvauspaikkoja, joita ei enää löytynyt kartalta samalla nimellä. Jos kuva oli otettu vuonna 1941 Kuusjärvellä, heräsi kysymys, täytetäänkö metatietoihin alkuperäinen kunta Kuusjärvi vai nykyinen kunta Outokumpu. Olin kuullut periaatteen, ettei alkuperäistä tietoa saa menettää, joten syötin kunnaksi alkuperäisen kuvauskunnan. Jotta kuvaa voisi hakea tietokannan kategorioiden avulla myös nykyisellä kuntanimellä, kehitin tässä kappaleessa esiteltävän logiikkapohjaisen menetelmän.

5.1 Geologian tutkimuskeskuksen valokuvatietokanta

Geologian tutkimuskeskuksen sisäisessä käytössä oleva valokuvatietokanta on toteutettu käyttäen Photools.com-yrityksen kehittämää IMatch-kuvatiedostojen hallintaohjelmaa. Tietokanta sisältää kaikki GTK:n digitaalisessa muodossa olevat kuvat, joita on kaiken kaikkiaan noin 36 000 kpl (6/2015). Läheskään kaikkiin kuviin ei ole vielä ehditty syöttämään metatietoa. Lisäksi maastohavaintokuvaan ei ole tarkoituksaan täyttää kaikkia metatietoja, vaan metatiedot liitetään niihin palvelusovelluksessa käyttäen avaimena havaintotunnusta, jonka mukaan kuvat nimetään.

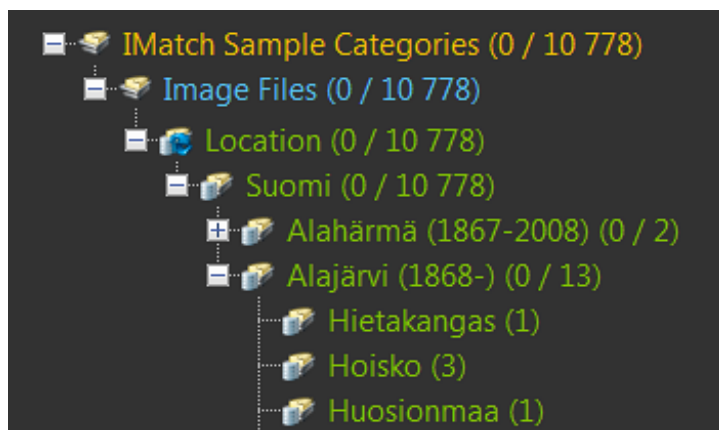
Aiemmin kuvien metatietojen tallentamiseen käytettiin IPTC-IIM skeemaa [12]. Vuoden 2015 alusta asti metatiedot on tallennettu Metadata Working Groupin¹⁴ (MWG) suosittelemiin ja ISO-standardoituuihin XMP-kenttiin [13]. XMP-metatiedot voidaan tallentaa sarjallistetussa RDF/XML-formaatissa kuvatiedostoihin upotettuna tai erillisiin XMP-tiedostoihin. GTK:n kuvatietokannassa käytetään kuvatiedostoihin upotettuja metatietoja, jolloin ne seuraavat aina kuvan mukana. Metatiedot sisältävistä kuvista 10 780 on otettu Suomen alueelta ja sisältävät tiedon kuvauskunnasta. Tässä työssä esitellyssä paikkojen muutostietojen mallintamisessa keskitytään niihin.

5.2 Muutosketjujen kuvaaminen dynaamisilla kategorioilla

IMatch-ohjelma luo ja ylläpitää kategorioita automaattisesti määriteltyjen metatietokenttien sisällön mukaan. Myös hierarkia on mahdollista ottaa mukaan mihin hyvänsä kategoriasisältöön käyttämällä sopivaa hierarkiaerotinta, esimerkiksi merkkiä “[]”. Paikannimille IMatch luo hierarkkiset kategoriat automaattisesti kuvatietoihin syötettyjen metatietokenttien valtio (XMP:Country), suuralue (XMP:State), kunta (XMP:City) ja paikka (XMP:Location) välille.

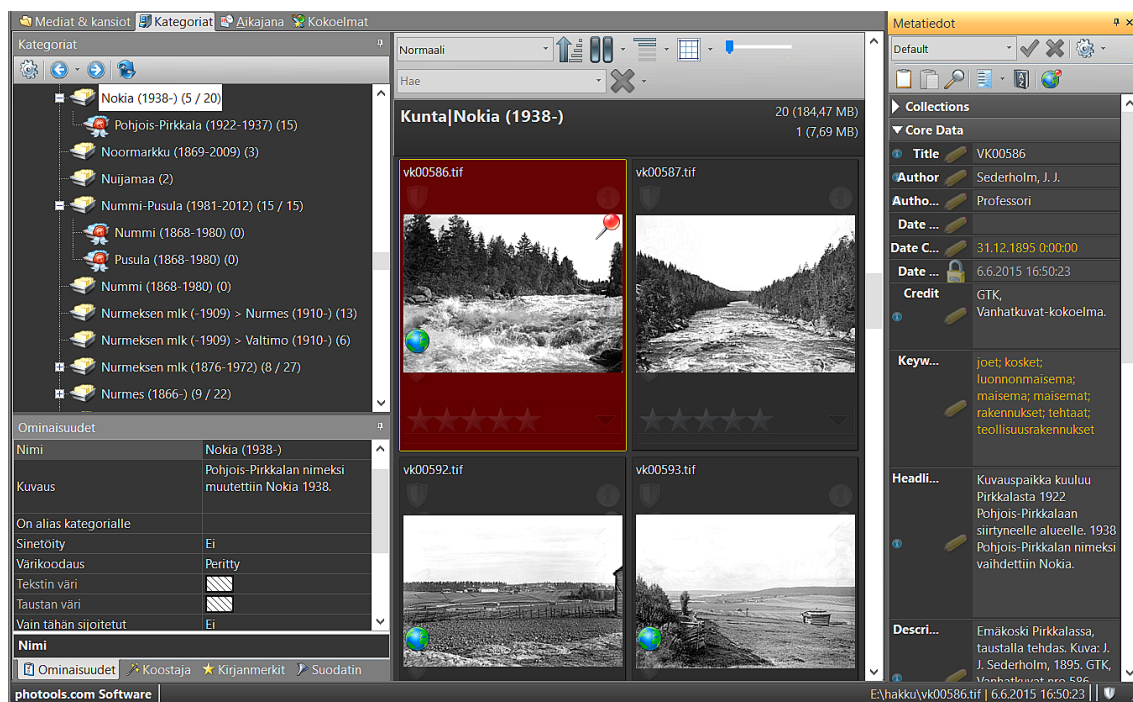
Kuvassa 39 on esimerkki näin syntyvästä hierarkiapuusta. Esimerkissä paikkatietokenttä suuralue on jätetty tyhjäksi ja se puuttuu siten myös paikkahierarkiasta. Suomessa suuralueita (lääni tai maakunta) käytetään varsin harvoin aineistojen metatiedoissa, toisin kuin esimerkiksi osavaltioita Yhdysvalloissa. Kaikki ajalliset kunnat sijoittuvat hierarkiatasoon kunta, joka on tässä tapauksessa valtion Suomi seuraava alempi taso.

¹⁴ <http://www.metadataworkinggroup.org/>



Kuva 39. IMatch-ohjelman automaattisesti luomat hierarkkiset paikkakategoriat

IMatch-ohjelmassa on myös mahdollista luoda kategorioiden välisiä suhteita loogisten operaattoreiden (AND, NOT, OR) avulla. Voidaan esimerkiksi määrittää, että kuva, joka kuuluu kuntakategoriaan Kuusjärvi, näkyy automaattisesti myös kategoriassa Outokumpu, joka on Kuusjärven nykyinen nimi. Kuvassa 40 nähdään IMatch-ohjelman käyttöliittymä. Ruudun vasemmassa reunassa on kategorianäkymä, keskellä näkyvät valitun kategorian sisältämät kuvat ja oikean reunan metatietoeditorilla syötetään kuvien metatiedot.



Kuva 40. IMatch-ohjelman käyttöliittymä.

GTK:ssa käytettyyn IMatch-ohjelmaan olen tehnyt asennuspaketin, jossa on jo valmiiksi syötetty kaikki Suomesta kokoamani ajalliset kunnat. Kuvatietojen tallentajan tarvitsee vain valita niistä oikea. Seuraavissa kappaleissa esitän muutosmuoto kerrallaan, miten muutokset hallitaan IMatchin kategorioissa, niin että kuva löytyy sekä alkuperäisen ottopaikan että nykyisen hallinnollisen paikan kategoriasta.

Perustaminen

Hallinnollisen alueen perustamisvuosi merkitään sen kategorian nimen perässä olevien sulkujen sisään ennen väliviivaa. Merkintätapa kuvaa sitä, että kyseisen hallinnollisen alueen olemassaolo alkoi tuona vuonna. Esimerkiksi Tampereen kaupunki perustettiin vuonna 1779. Silloin kunnaksi merkitään Tampere (1779-). Se, ettei väliviivan jälkeen ole vuosilukua, tarkoittaa että Tampere on edelleen olemassa.

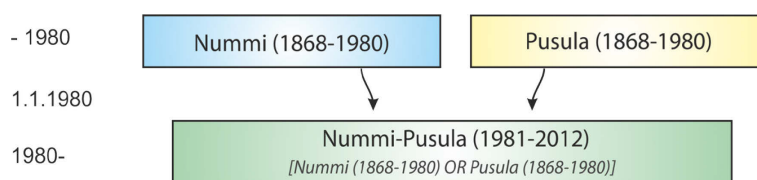
Lakkauttaminen

Kyseisen nimisen alueen olemassaolon päättymisen ilmoitetaan nimen jälkeen olevissa suluissa väliviivan jälkeisen päättymisvuoden avulla. Päättymisen voi johtua paitsi hallinnollisen alueen lakkauttamisesta, myös sen jakaantumisesta erinimisiksi hallinnollisiksi alueiksi, yhdistymisestä toiseen alueeseen, nimenmuutoksesta tai luovuttamisesta toiseen valtioon. Kategorian kuvailukentässä kerrotaan tarkemmin päättymisen syy, päiväys ja mahdolliset hallinnolliset alueet, mihin kyseiset alueet myöhemmin sijoittuvat.

Yhdistyminen

Jos kaksi tai useampi hallinnollinen alue yhdistyvät uuden nimiseksi hallinnolliseksi alueeksi, uudelle alueelle luodaan kategoria, johon yhdistyvien kategorioiden kuvat kytetään OR-parametrilla. Kaikki näin kytketyissä kategorioissa olevat kuvat näkyvät myös uudessa kategoriassa. Jos yhdistyminen tapahtuu johonkin jo olemassa olevaan hallinnolliseen alueeseen, toimintansa lopettavat hallinnolliset alueet kytetään siihen, eikä tarvitse luoda uutta kategoriaa. Jos haluaisimme tarkemmin erottaa eriaikaisissa alueissa sijainneet kuvat, voisimme luoda uuden kategorian aina uuden aluemuutoksen tapahtuessa. Näin ketjutetuista kategorioista voisimme katsoa kuvia tarkalleen tietyistä ajallisesta paikasta. Käytännössä tämä on osoittautunut epäkäytännölliseksi, koska kategorioiden määrä kasvaisi tällä menetelmällä hyvin suureksi. Siksi kustakin samannimisenä pysyneestä hallinnollisesta alueesta käytetään vain yhtä kategoriaa.

Nummen ja Pusulan kuntien yhdistyminen Nummi-Pusulaksi mallinnetaan sijoittamalla kategoriat yhdistävä kaava Nummi-Pusulan kategoriaan. Yhdistymistä esittävässä kuvassa 41 kaava on kirjoitettu kursivoilla.



Kuva 41. Nummen ja Pusulan kuntien yhdistäminen Nummi-Pusulaksi vuoden 1981 alussa esitettynä dynaamisten kategorioiden avulla.

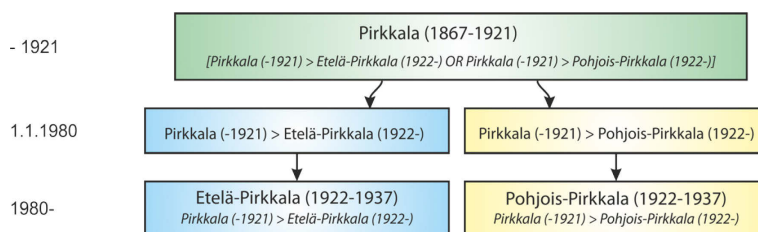
Jakaantuminen

Kuten kuvassa 26 näytettiin, on tarpeen käyttää tarkempaa paikkatietoa, jotta jakaantuneelta alueelta aiemmin otettu valokuva pystytään kategorisoimaan parhaalla mahdolli-

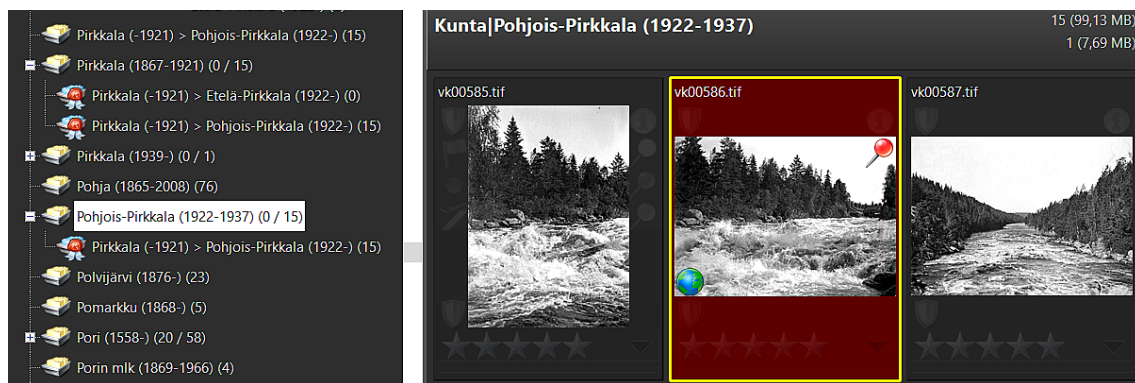
sella tavalla. Jakaantuneiden alueiden mallintamisessa käytetään erillisiä muutuskategorioita kaikille jakaantumisen tuloksena oleville alueilla. Kuvat sijoitetaan näihin muutuskategorioihin, mistä ne automaattisesti kopioituvat myös muihin tarvittaviin kategorioihin. Kuvaa kategorisoitaessa olisi siis hyvä tietää, mihin hallinnolliseen alueeseen se on myöhemmin kuulunut. Näin se voidaan sijoittaa oikeisiin muutuskategorioihin. Jollei kuvan ottopaikan alkuperäistä kuntaa tarkempi sijainti ole tiedossa, yhteys myöhempiin ajallisiin paikkoihin katkeaa.

Koska muutokset tapahtuvat yleensä (93 prosentissa tapauksista) vuodenvaihteessa [55], muutuskategorioiden päiväyksenä käytetään muutosta edeltävää ja muutoksen tapahtumavuotta. Alueiden siirtymisen suuntaa kuvataan kategorian nimessä olevalla nuolella. Esimerkiksi Pirkkalan kunnan jakaantuminen vuonna 1922 mallinnetaan kategorioiden avulla kuvassa 42 esitetyllä tavalla.

Käytännössä kuvat sijoitetaan muutuskategoriaan Pirkkala (-1921) > Pohjois-Pirkkala (1922-). Kategoriakaavojen ansiosta kuvat näkyvät tällöin myös alkuperäisen kuvauskunnan Pirkkala (1867-1921) ja myöhemmän kuvauskunnan Pohjois-Pirkkala (1922-1937) kategorioissa, kuten kuvassa 43.



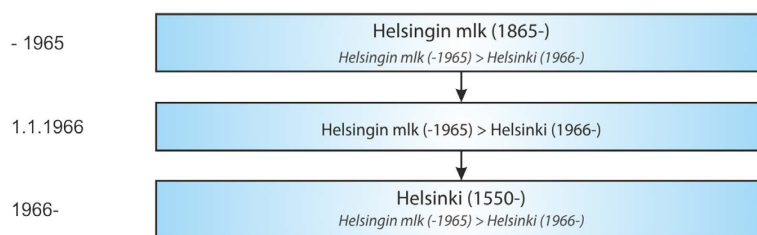
Kuva 42. Pirkkalan jakaantuminen Etelä-Pirkkalaksi ja Pohjois-Pirkkalaksi vuoden 1922 alussa esitettyinä dynaamisten kategorioiden avulla.



Kuva 43. Pirkkalan kunnassa otettu kuva on sijoitettu alun perin kategoriaan Pirkkala (-1921) > Pohjois-Pirkkala (1922-). Kaavan ansiosta se näkyy myös kategoriassa Pohjois-Pirkkala (1922-1937)

Aluesiirto

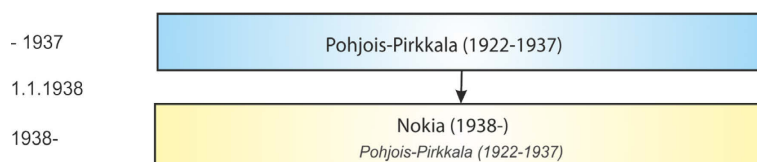
Siirtyvälle alueelle luodaan oma muutuskategoria, joka päivätään siirtovuoden mukaan. Kategorian kuvailukentässä kerrotaan tarkemmin, mitä alueita on siirtynyt, esimerkiksi kylien nimien avulla. Siirtyneiltä alueilta otetut kuvat pystytään näiden tietojen perusteella sijoittamaan oikeaan kategoriaan. Kuvassa 44 on esimerkki vuoden 1966 alussa tapahtuneesta aluesiirrosta.



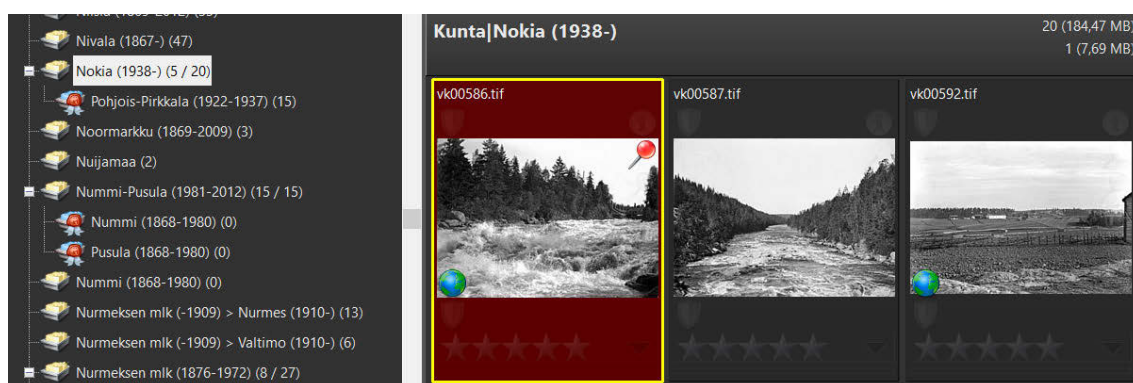
Kuva 44. Helsingin maalauskunnasta siirrettiin Vuosaaren alue Helsingin kaupunkiin 1966.

Nimenmuutos

Hallinnollisen alueen vanhan nimen mukaan nimetystä kategoriasta löytyvät kaikki kuvat, jotka on otettu kunnassa ennen nimenmuutosta. Uudella kunnanimellä nimetystä kategoriasta löytyvät sekä vanhalla että uudella nimellä merkityt kuvat. Kuvassa 45 on esimerkki kuinka kategoriat on toteutettu. Kuva 46 näyttää miten alun perin Pirkkalan kunnassa otettu kuva näkyy myöhemmässä Pohjois-Pirkkala (1922-1937) kategoriassa ja nimenmuutoksen jälkeisessä Nokia (1938-) kategoriassa.



Kuva 45. Pohjois-Pirkkalan nimeksi muutettiin Nokia vuoden 1938 vaihteessa.



Kuva 46. Pohjois-Pirkkalan nimenmuutos Nokiaksi näkyy IMatchin kategorioissa niin että kaikki Pohjois-Pirkkalan kategoriassa olevat kuvat sijoittuvat myös Nokian kategoriaan.

Alueluovutus

Suomen vuoden 1940 kokonaisia kuntia käsittävät alueluovutukset Neuvostoliitolle merkitään hallinnollisen alueen lakkaamisena vuonna 1940. Kuvailukentässä kerrotaan loppumisen syyksi alueluovutus.

Osittain luovutettu hallinnollinen alue käsitellään samalla tavalla kuin jakaantuminen. Kuvassa 47 on esimerkkinä Vahvialan kunnan osittainen alueluovutus vuonna 1940 ja sen Suomen puolelle jääneen alueen jakaminen Lappeen, Lappeenrannan ja Ylämaan kuntiin vuoden 1946 alussa. Kaikki muutuskategorioihin sijoitetut kuvat näkyvät kategoriassa Vahviala (1921-1945), johon sijoitetaan myös sellaiset kuvat, joiden tarkempi kuvauspaikka ei ole jostain syystä tiedossa.

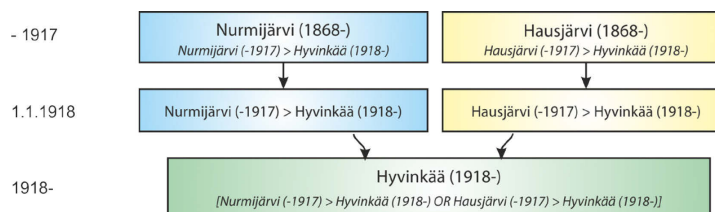


Kuva 47. Suurin osa Vahvialan kunnasta jäi luovutetuille alueille. Loput alueet jaettiin kolmen kunnan kesken. Katteoria Vahviala (1921–1945) kuvaa koko Vahvialan kunnan olemassaoloaika ja toimii koostekategoriana kaikille Vahvialassa otetuille kuville.

Aluesiirtoperustaminen

Aluesiirtoperustaminen käsitellään luomalla perustettavaan kuntaan siirtyville alueille muutuskategoriat, joihin siirtyviltä alueilta otetut kuvat sijoitetaan. Muutuskategoriat kopioivat kuvat haluttuihin ajallisiin kuntakategorioihin. Kuva 48 esittää Hyvinkään kunnan perustamista vuoden 1918 alussa Nurmijärven ja Hausjärven kunnista siirretyille alueille.

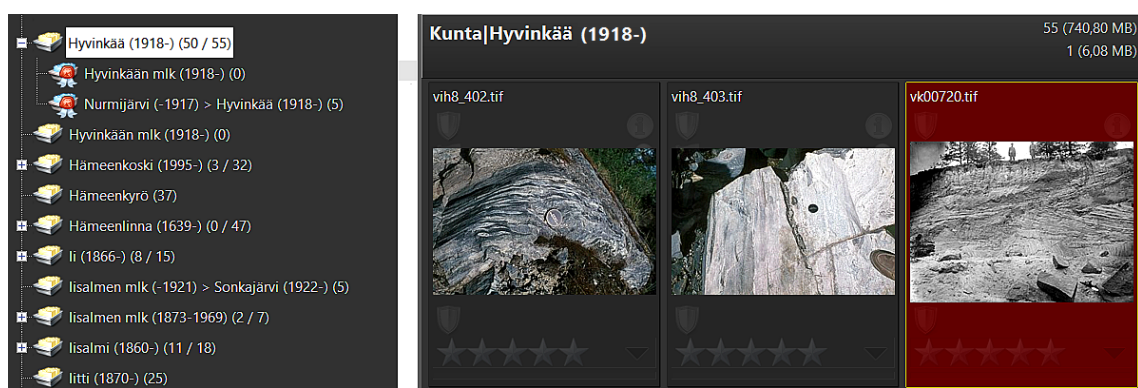
Kuvassa 49 Nurmijärven kunnassa Hyvinkään aseman lähellä otettu kuva näkyy kategoriassa Nurmijärvi (1868-). Alun perin se on sijoitettu muutuskategoriaan Nurmijärvi (-1917) > Hyvinkää (1918-). Muutuskategoria kopioi kuvan myös myöhempään ajalliseen kategoriaan Hyvinkää (1918-), kuten kuvasta 50 nähdään.



Kuva 48. Hyvinkään kunta perustettiin vuonna 1918 Nurmijärvestä ja Hausjärvestä siirretyille alueille.



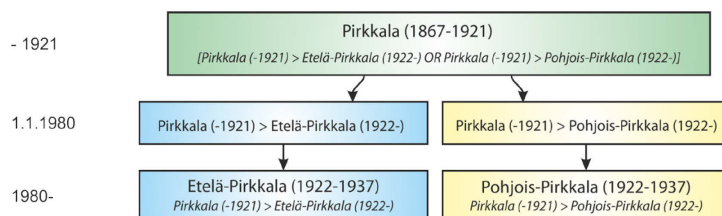
Kuva 49. Nurmijärvellä vuonna 1895 otettu valokuva, jonka kuvauspaikka sijaitsee kunnasta Hyvinkääseen vuonna 1918 siirtyneellä alueella.



Kuva 50. Edellisessä kuvassa esitetty harjuleikkauksen valokuva näkyy myös kategoriassa Hyvinkää (1918-).

Aluesiirtojakaantuminen

Aluesiirtojakaantumisessa hallinnollinen alue jakaantuu useamman kunnan kesken. Kuvan 51 näyttämässä esimerkissä muutuskategorioihin sijoitetaan Bromarvin alueilla otetut kuvat sen mukaan, sijoittuvatko kuvat jakaantumisen jälkeen Tenholaan vai Hankoon.



Kuva 51. Bromarvin kunnan alueet jaettiin Hangon ja Tenholan kesken 1.1.1977.

5.3 Käytännön kokemuksia kategoriamenetelmän toimivuudesta

Kategoriamenetelmä muuttuvien kuntatietojen hallinnassa on ollut käytössä jo yli kymmenen vuoden ajan Geologian tutkimuskeskuksessa. Tässä kappaleessa esitän käytännön kokemuksia sen käytöstä. Kuvien paikkatietojen syöttäminen menetelmän avulla on ollut yleensä ottaen helppoa, koska käytettävissä olevien ajallisten kuntien luettelo on ollut valmiina. Samasta syystä myös kuviin aiemmin syötetyissä metatiedossa olleita virheitä on voitu korjata. Tyypillinen virhe on se, että kuvan ottopaikaksi on alkuperäisen sijasta merkitty myöhempi alueella sijainnut kunta. Siinä tapauksessa alkuperäinen sijaintitieto on voitu menettää ja paikkatiedon tarkkuus huonontua. Vanhempien kuvien paikkatiedot on onneksi voitu tarvittaessa tarkistaa alkuperäisestä kuvaluettelosta tai negatiivilehdestä. Eniten työtä ovat vaatineet paikat, jotka ovat kuvan ottohetken jälkeen jakaantuneet. Aina ei ole ollut helppoa paikantaa kuvan ottopaikkaa riittävällä tarkkuudella oikean muutoskategorian valintaa varten.

Myös kuvien hakeminen kuntakategorioiden avulla on toiminut hyvin. Sen sijaan tekstipohjainen haku voimassa olevalla kuntajaolla ei onnistu, sillä metatiedossa on vain kuvauspaikan alkuperäinen kuntatieto eikä haku kohdistu kategorioiden nimiin. Jonkin verran ongelmallista on myös se, että metatietoihin merkitty kunnan nimi vuosilukuihin ja mahdollisine nuolineen saattaa herättää ihmetystä sellaisessa kuvan käyttäjässä, joka ei tiedä sen tarkoitusta. GTK:n ulkopuoliseen Hakku-hakupalveluun viedyistä kuvista vuosiluvut on poistettu. Seuraavassa kappaleessa kuvataan sitä, miten kuntamuu-
tokset on siellä otettu huomioon.

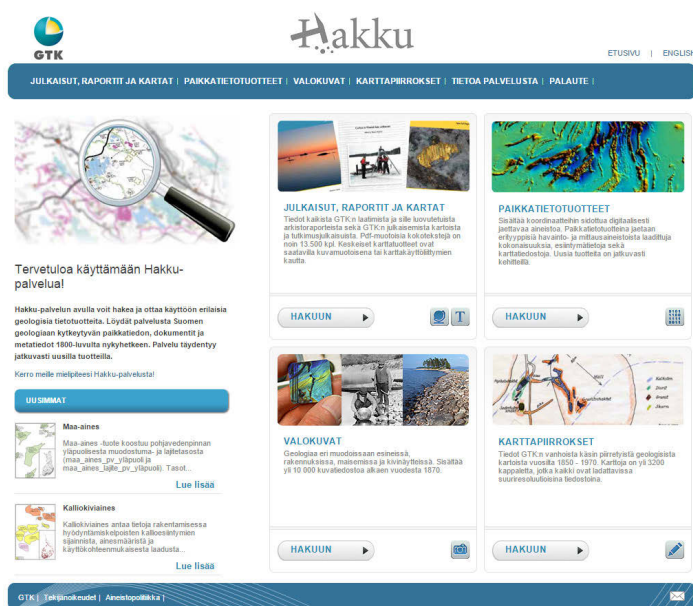
6 Muutostietojen mallintaminen verkkopalvelussa

6.1 Hakku-hakupalvelu

Geologian tutkimuskeskus julkaisi kokoavan Hakku-palvelun¹⁵ keväällä 2014. Olin mukana projektiryhmässä tekemässä palvelun määrittelyä ja toteutusvaiheessa testaa-massa ja tekemässä kehitysehdotuksia. Hakku-palvelusta voidaan hakea ja ladata GTK:n aineistoja neljästä osiosta:

- Julkaisut, raportit ja kartat
- Paikkatietotuotteet
- Valokuvat
- Karttapiirokset

Hakku-käyttöliittymä toteutettiin IWA Labsissa käyttäen Ruby on rails -verkkosovelluskehystä. Palvelu käyttää GTK:n itse kehittämää ja ylläpitämää TUTTI-tuotetietokantaa, joka on toteutettu Oracle-relaatiotietokantana. Katsottavat ja ladattavat aineistot on tallennettu TUPA-tuotepalvelimelle, joka replikoituu kerran vuorokaudessa sisäverkosta ulkoverkkoon.



Kuva 52. Hakku-hakupalvelun etusivu, jossa näkyy palvelun neljä hakuosiota.

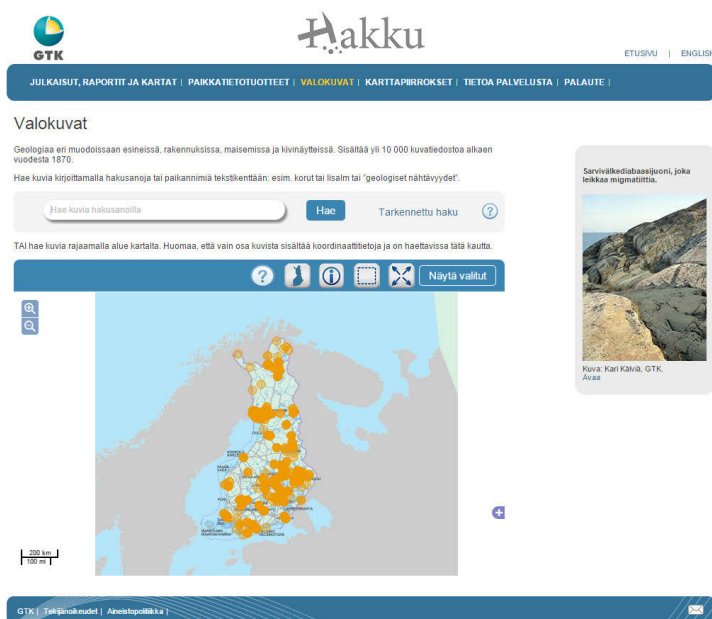
6.2 Valokuvat Hakku-palvelussa

Paikkatietoa Hakun valokuvuihin on tallennettu sekä paikannimien että koordinaattien avulla. Paikannimet on tallennettu tietokannan kenttiin: kuvausmaa, suuralue, kuvauskunta ja kuvauspaikka. Käyttöliittymässä paikannimet näytetään yhdessä kentässä vas-taavassa järjestyksessä suurimmasta alueesta pienimpään. Koordinaatit on tallennettu,

¹⁵ <http://hakku.gtk.fi/>

mikäli ne ovat olleet tiedossa, Suomen alueelta sekä WGS84 että ETRS-TM35FIN -koordinaatistoissa ja muualta maailmasta vain WGS84 koordinaatistossa. Koordinaattitieto on 2188 kuvassa eli noin viidesosassa palvelun kaikista kuvista (11/2015).

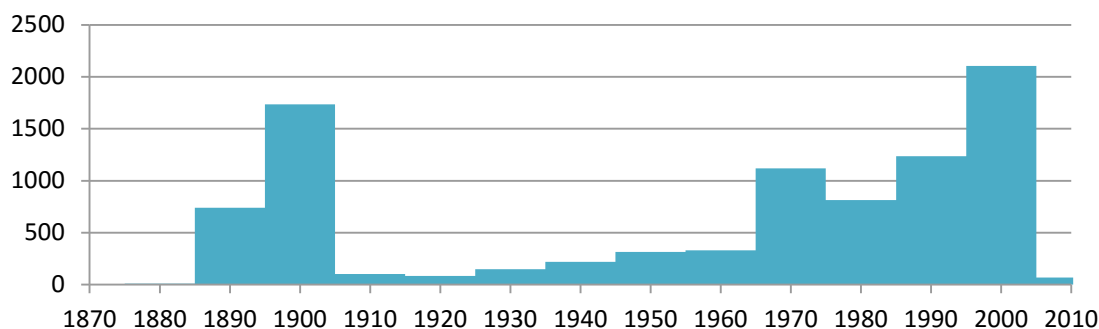
Käyttöliittymässä käytetään hakuun pelkästään ETRS-TM35FIN -koordinaatistoa ja kartalta voidaan siksi hakea kuvia vain nykyisen Suomen alueelta. Tämä rajoitus johtuu siitä, että valokuvaosiossa käytetään samaa karttakehystä kuin Hakku-palvelun paikkatieto-osiossa, jossa on haettavana vain Suomea koskevaa aineistoa ETRS -TM35FIN -koordinaatistossa. Kuvassa 53 nähdään Hakun valokuvahaun käyttöliittymä. Hakukenttään syötettyjä termejä haetaan kuvan asiasanoista, vapaamuotoisesta kuvatekstistä ja paikannimistä. Kuvia voidaan hakea myös rajaamalla alue karttakäyttöliittymästä kuviin syötettyjen koordinaattitietojen perusteella.



Kuva 53. Hakku-palvelun Valokuvat-osion etusivu.

Hakun valokuvaosiossa on 11300 kuvaa, joista 9724:ssä on merkitty kuntatieto (11/2015). Suomen alueelta otettuja kuntatiedon sisältäviä kuvia Hakussa on 9292 kpl. Tarkennetussa haussa paikannimihakua varten on oma hakukenttensä, johon syötettyjä termejä haetaan vain kuvan paikkatiedoista.

Ajallisesti Hakussa on kuvia 1870-luvulta tähän päivään. Vanhoja lasilevynegatiiveja 1900-luvun vaihteesta on skannattu järjestelmällisesti niiden historiallisen arvon vuoksi. Ne näkyvät kohoumana kuvien ajallisessa jakaumassa (Kuva 54). Muuten kuvien määrä kasvaa varsin tasaisesti 2000-luvulle saakka, mutta aivan uusimpia kuvia ei ole vielä ehditty viemään tietokantaan.



Kuva 54. Hakku-palvelussa olevien Suomessa otettujen ja kuntatiedon sisältävien kuvien ajallinen jakauma.

6.3 Kuntamuutokset Hakku-palvelussa

Hakun Suomen alueelta otetuista 9292 kuvasta 2129 eli 23 % on otettu sellaisella alueella, joka on liitetty toiseen kuntaan, toiseen valtioon tai vaihtanut nimeä kuvanottohetken jälkeen. Jos haku tehdään nykyisen kunnannimistön perusteella, nämä kuvat eivät löytyisi tietokannasta normaaleilla hakutyökaluilla.

Olen kehittänyt Hakku-palveluun yksinkertaisen menetelmän, jonka avulla hakuja voidaan tehdä käyttäjän valinnan mukaan joko alkuperäisen tai nykyisin käytössä olevan kuntajaon perusteella. Yleensä aineistoja pystyy palveluista hakemaan jomman kumman tai molempien mukaan riippuen siitä miten aineistot on kuvailtu. Hakku on tietämäni mukaan ensimmäinen palvelu ainakin Suomessa, jossa käyttäjä voi itse valita, kumpaa jakoa käyttää. Tätä työtä varten tekemäni kyselyn mukaan molempia hakutapoja tarvitaan erilaisia käyttäjiä varten. Historiaa tuntevat hakijat voivat haluta kohdistaa hakunsa juuri tietyn ajan kohteisiin tietäen haluamansa ajan kuntajaon. Monet muut käyttäjät eivät välttämättä tiedä kuntamuutoksista ja haluavat hakea aineistoa parhaiten tuntemiensa kuntien mukaan, jotka ovat yleensä olemassa olevia.

Menetelmän avulla pystytään mallintamaan kuntaliitokset ja kuntien nimenmuutokset. Sen sijaan kuvanottohetken jälkeen jakaantuneella alueella otetut kuvat eivät löydy, jos ne sijaitsevat erinimisessä kunnassa kuin alun perin. Onneksaasti monet kuntien jakaantumiset ovat kompensoituneet myöhemmin, esimerkiksi kun alun perin maalaiskunnasta eronneet taajama-alueet ovat myöhemmin imaisseet alkuperäisen emokuntansa itseensä. Esimerkiksi vuonna 1929 tapahtunut Rovaniemen jakaantuminen kauppalaan ja maalaiskuntaan kompensoitui 2006, kun Rovaniemen mlk liitettiin Rovaniemen kaupunkiin. Samalla kompensoituivat monet aikaisemmat aluesiirrot, joita kuntien välillä oli tapahtunut Rovaniemen kauppalan laajetessa vaihteittain maalaiskunnan alueelle.

Kehittämäni menetelmä on tällä hetkellä käytössä Hakun valokuvaosiossa. Se voidaan ottaa myöhemmin käyttöön myös muissa osioissa lukuun ottamatta paikkatietotuotteita, joiden paikkatiedot perustuvat koordinaatteihin eivätkä paikannimistöön. Käyttöönottoa muissa osioissa on hidastanut mm. niissä käytetyn paikkatiedon luokittelun puuttuminen eli kunnannimet on syötetty samaan kenttään muiden paikannimien kanssa, jolloin niiden laajentaminen on hankalampaa.

Menetelmän toiminta pohjautuu siihen, että Hakku-hakupalvelun käyttämään tuotetietokantaan (TUTTI) on lisätty kuntamuutostaulu, joka sisältää tiedot kuntien nimenmuutoksista ja kuntaliitoksista. Kuvassa 55 on ote muutostaulun sisällöstä. Jos käyttäjä tekee oletusasetuksilla haun "Akaa", tulokseen tulevat mukaan myös Kylmäkoskessa

otetut kuvat, koska entinen Kylmäkosken kunta on nykyään osa Akaan kuntaa. Jos haki- ja haluaa löytää vain ne kuvat, jotka on otettu "Akaa"-nimisessä kunnassa, hän voi rajoittaa käyttöliittymässä olevan kohdan: "Hae vain valitulla kunnan nimellä".

Kuntamuutos_ID	Mihin_kuntaan	Kunta_ID	Kunta_nimi
1	Akaa	7	Toijala
2	Akaa	7	Viiala
3	Akaa	7	Kylmäkoski
4	Alajärvi	1	Lehtimäki
5	Alavus	3	Töysä
6	Eura	13	Hinnerjoki
7	Eura	13	Honkilahti
8	Eura	13	Kiukainen
9	Eura	13	Köyliö
10	Eura	13	Säkylä
11	Hamina	24	Vehkalahti
12	Hanko	27	Bromarv
13	Hattula	30	Tyrväntö

Kuva 55. Ote Tutti-tietokannan taulusta, johon on tallennettu tiedot kuntamuutoksista

Kuvassa 56 esitetään Hakun käyttöliittymän kunnannimen hakulaajennoksen valintaruutu ja siihen liittyvä ohjeteksti aktivoituna. Ohjeteksti opastaa käyttäjää, miten paikannimihaun laajennus toimii. Se on tärkeää, jotta käyttäjä ymmärtää, miksi hakutuloksessa voi oletusasetuksella olla kuvia erinimisistä kunnista, kuin millä haku on alun perin tehty.



Kuva 56. Hakku-palvelun valokuvaosion tarkennettu haku, jossa on tehty haku Valkeakosken kunnasta ilman laajennusta muihin alueella sijainneisiin kunnannimiin.

Menetelmä huolehtii vain sellaisten kuntamuutosten mallintamisesta, joissa muutos koskee koko kunnan aluetta. Sellaisia ovat kuntien yhdistymiset ja nimenmuutokset. Tällöin haun tarkkuus ei huonone lainkaan saannin kuitenkin parantuessa.

Menetelmän toimivuuden kannalta on hyvä, että jakaantumiset sijoittuvat Suomessa ajallisesti 1900-luvun alkupuolelle, kuten kuvasta 24 nähtiin. Kuntien yhdistymiset sen sijaan sijoittuvat 1900-luvun loppupuolelle ja 2000-luvun alkuun. Jakaantuminen on tapahtunut niin kauan sitten, ettei se vaikuta hakutuloksen saantiin kuin kaikkein vanhimpien kuvien osalta, tai myöhemmin tapahtunut yhdistyminen on kompensoinut jakaantumisen vaikutuksen. Menetelmän käytöstä ei olisi mitään haittaa, vaikka sitä käytettäisiin sellaisessakin tilanteessa, että jakaantumisten ja yhdistymisten ajallinen jakau-

tuminen olisi toisenlainen. Menetelmällä saavutettava hyöty tosin olisi sellaisessa tilanteessa pienempi.

Jakaantumiset voitaisiin käsitellä myös niin, että jakaantuneeseen kuntaan tehty haku laajennettaisiin kaikkiin myöhemmin alueella sijaitseviin kuntiin. Esimerkiksi haettaessa kunnalla Nokia, saisimme tuloksena myös kaikki kuvat vanhasta Pirkkalan kunnasta (1867–1921). Tällöin haun saanti paranisi, mutta tarkkuus huononisi, koska vain osa Pirkkalassa otetuista kuvista on otettu nykyisen Nokian alueella. Tätä ei ole toteutettu, koska on haluttu välttää virheellisiä hakutuloksia.

7 Tulokset

Tarkkuus- ja saantiarvoja käytetään yleisesti hakusovellusten toimivuuden mittareina. Hakusovellus pyritään optimoimaan niin, että saanti olisi mahdollisimman hyvä ilman, että tarkkuus putoaisi häiritsevästi [32]. Kun halutaan selvittää, parantaako käytetty menetelmä hakutulosten tarkkuutta ja saantia, kannattaa mitata arvot sekä menetelmän ollessa käytössä että ilman sitä. Tulosten erotus kertoo menetelmän toimivuudesta. Hyvin erityyppisiä menetelmiä ei voi aina sellaisenaan verrata keskenään, koska aineistolliset lähtökohdat ja menetelmien käyttötavat voivat poiketa liikaa toisistaan. Siksi tässä työssä ei oteta varsinaisesti kantaa menetelmien väliseen keskinäiseen paremmuuteen.

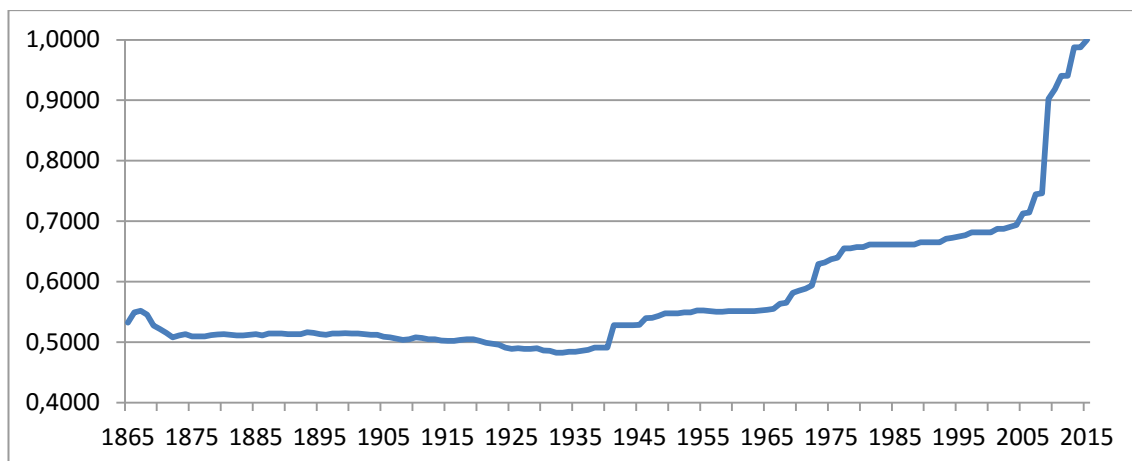
Olen kartoittanut kyselyn avulla suomalaisten muisti- ja tutkimusorganisaatioiden valmiutta ja tarvetta ajallisen paikkatiedon hallintaan. Erityisesti olen kiinnostunut siitä, hakevatko organisaatioiden tietojärjestelmien käyttäjät mieluiten alkuperäisillä vanhoilla kuntanimillä vai nykyisin käytössä olevilla kunnilla. Esittelen yhteenvedon kyselyyn saamistani vastauksista tämän kappaleen viimeisessä osiossa.

7.1 Tarkkuus- ja saanti ilman muutostenhallintaa

Tässä kappaleessa lasketaan tarkkuuden ja saannin perustaso tilanteessa, jossa hallinnallisten alueiden muutostenhallinnassa ei ole käytössä mitään menetelmää. Tutkimuksessa oletetaan, että aineistoa haetaan kuntien perusteella ja haussa käytetään nykyistä kuntanimistöä. Tämänäyttöisessä laskennassa käytetään tyypillisesti jotain esimerkkiaineistoa, esimerkiksi Kauppinen ja muut [23] käyttivät kahta erilaista kuvajoukkoa laskiesseen, miten SAPO-ontologian käyttö vaikutti tehtyjen hakujen tarkkuus- ja saantiarvoihin oletuksella, että aineisto jakaantuu tasaisesti tutkittavien kuntien alueille.

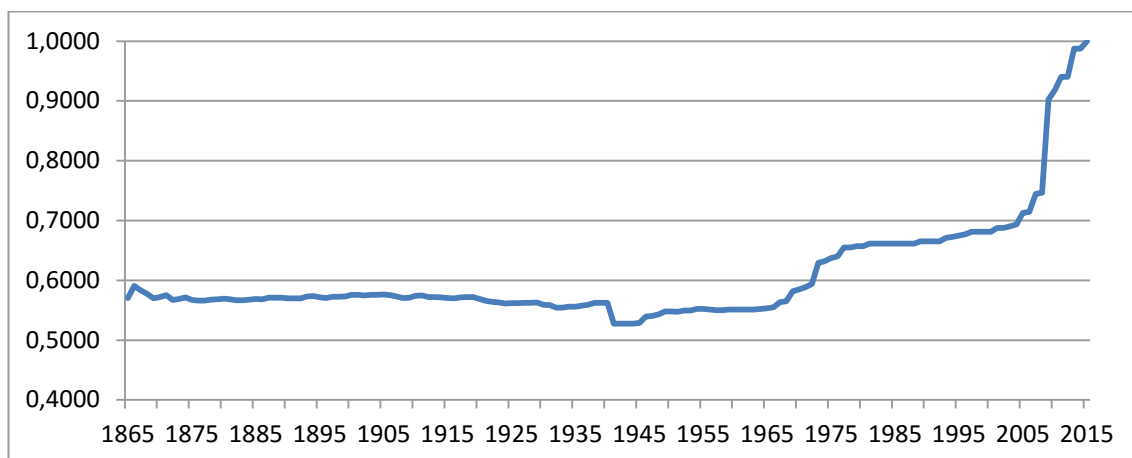
Minua kiinnosti tietää, miten Suomen hallinnollisissa alueissa tapahtuneet muutokset vaikuttavat hakutuloksiin ajan suhteen muuttuvalla aineistolla. Siksi tein vertailun sijoittamalla yhden kuvitteellisen kohteen jokaiseen Suomen kuntaan jokaiselle vuodelle alkaen vuodesta 1865, jolloin Suomeen perustettiin ensimmäiset kunnat. Kohteiden vuosittaiset määrät ovat siis samoja kuin Suomen kuntien lukumäärät, jotka nähtiin kuvassa 2. Tällä asetelmalla olen laskenut, löytyvätkö kohteet haettaessa niitä nykyisillä kuntanimillä. Tulos ei tällöin ole riippuvainen jonkin esimerkkiaineiston kohteiden ajallisesta jakaantumisesta.

Kuvassa 57 on saamani tulos. Pystyakselilla on löydettyjen kohteiden (eli muuttumattomina pysyneiden) kuntien määrä jaettuna kaikkien kyseisen vuoden kuntien määrällä. Hakutuloksessa relevantiksi katsotaan kohde, joka sijaitsee edelleen tällä hetkellä samannimisessä kunnassa, johon se on alun perin sijoitettu. Tällöin pystyakselin arvo on sama kuin hakutuloksen saanti tarkkuuden ollessa koko ajan yksi eli kaikkien tulosten ollessa relevantteja. Esimerkiksi hakiessamme tämänvuotisia kohteita tämänvuotisella kuntajaolla saanti on yksi eli 100 % mutta hakiessamme vuoden 1915 kohteita saanti on 0,50 eli 50 %. Toisin sanoen tuhannesta vuoden 1915 kohteesta löytyisi 500 kohdetta. Vuoden 2009 kohdalla oleva jyrkkä muutos käyrässä johtuu tuona vuonna tapahtuneesta suuresta kuntaliitosten määrästä. Vuoden 2009 alussa 67 kuntaa lakkasi olemasta yhdistyessään toiseen kuntaan.



Kuva 57. Hakutulosten saannin perustaso olettaen, että kohteet jakaantuvat tasaisesti Suomen kaikkien kuntien kesken. Kaikki saadut tulokset ovat relevantteja eli tarkkuus on koko ajan 100 %.

Jos huomioidaan kappaleessa 3.2 esitetty hypoteesi, että luovutetun alueen kuntia haetaan Suomessa lähes yksinomaan luovutusajankohdan mukaisella eikä nykyisellä venäläisellä kuntajaolla, saadaan alueluovutuksia edeltävälle ajalle äskeitä parempi tulos (Kuva 58). Nyt tuhannesta vuoden 1915 kohteesta löytyy 570 kappaletta.



Kuva 58. Hakutulosten saannin perustaso, kun oletetaan, että luovutetun alueen kohteita haetaan luovutusajan kuntajaolla.

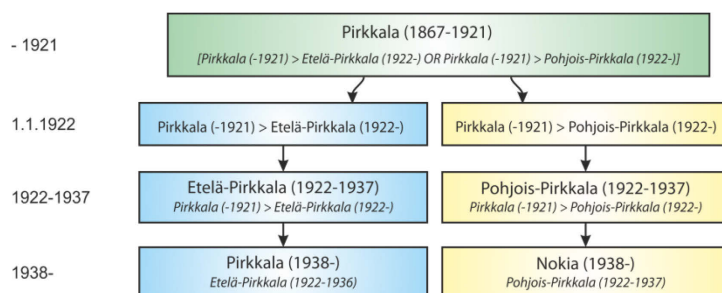
7.2 Tarkkuus- ja saanti kategoriamenetelmää käytettäessä

Geologian tutkimuskeskukselle toteuttamani kategoriamenetelmä esiteltiin kappaleessa viisi. Menetelmä on käytettävissä vain IMatch-aineistohallintaohjelmassa, mikä estää sen laajemman soveltamisen sellaisenaan. Sen käyttämää loogista periaatetta on kuitenkin mahdollista soveltaa myös muualla. Kategoriamenetelmässä haku ei tapahdu syötämällä hakukenttään asiasanoja vaan valitsemalla aakkostetusta luettelosta haluttu kategoria.

IMatch-ohjelman loogiset kategoriat mahdollistavat eräänlaisten ajallisten kategoriaketjujen toteuttamisen. Nuo kategoriaketjut kuljettavat aineiston ajallisten kuntakategorioiden läpi aina uusimpiin voimassaoleviin kuntiin saakka. Esimerkkinä pidemmistä

kategoriaketjuista kuvassa 59 nähdään Pirkkalan kunnan jakaantuminen, sitä seuraavat nimenmuutokset. Pirkkalan kunnan pohjoisosasta vuonna 1905 otettu valokuva sijoitetaan kategoriaan *Pirkkala (-1921) > Pohjois-Pirkkala (1922-)*. Kategoriaketjujen avulla se näkyy sekä vanhassa Pirkkalassa että nykyisessä Nokiassa. Mikäli ei ole tiedossa, mistä osasta Pirkkalaa aineisto on peräisin, se voidaan sijoittaa vain Pirkkalan kategoriaan, jolloin se ei löydy myöhemmistä kategorioista.

Vastaavalla tavalla voidaan luoda ketjuja yli kaikkien muutosmuotojen, jopa lukuisien pienten aluesiirtojen. Käytännön syistä kaikkia aluesiirtoja ei ole kuitenkaan ketjutettu valmiiksi läpi historian, koska kategorioiden määrä kasvaisi tällöin hyvin suureksi. Ketjutuksia voidaanakin tehdä tarpeen mukaan vain sellaisille alueille, joilta on olemassa aineistoa.



Kuva 59. Kategorioiden ketjuja, joiden avulla aineisto siirtyy myöhemmin sijainneisiin kuntiin.

Muutuskategorioiden käyttö mahdollistaa jokaisen pienenkin siirtyvän alueen käsittelemisen yksilöllisesti. Tiedonhaun saannin ja tarkkuuden kannalta kategoriamenetelmä mahdollistaa siksi teoriassa sekä täyden tarkkuuden että saannin. Edellytyksenä sille on, että ketjutus on kattava ja luokittelija osaa sijoittaa kuvat alun perin oikeisiin kategorioihin. Se vaatii yleensä kuntatietoa tarkemman paikkatiedon olemassaoloa ja käyttöä. Muussa tapauksessa tarkkuus pysyy aina täytenä, mutta saanti putoaa sen mukaan, miten hyvin aineisto pystytään sijoittamaan oikeisiin kategorioihin käytettävissä olevan paikkatiedon perusteella.

7.3 Tarkkuus- ja saanti tietokannan muutostaulua käytettäessä

Hakku-hakupalvelussa käytetty muutostaulumenetelmä esiteltiin kappaleessa kuusi. Menetelmän hyvän toimivuuden avaintekijä on se, että hallinnollisten alueiden yhdistymiset ovat tapahtuneet myöhemmin kuin jakaantumiset. Menetelmä on toteutettu niin, että virheellisiä hakutuloksia ei saada, vaan haun tarkkuus on aina 100 %. Saanti sen sijaan voi laskea jakaantumisten ja aluesiirtojakaantumisten vuoksi.

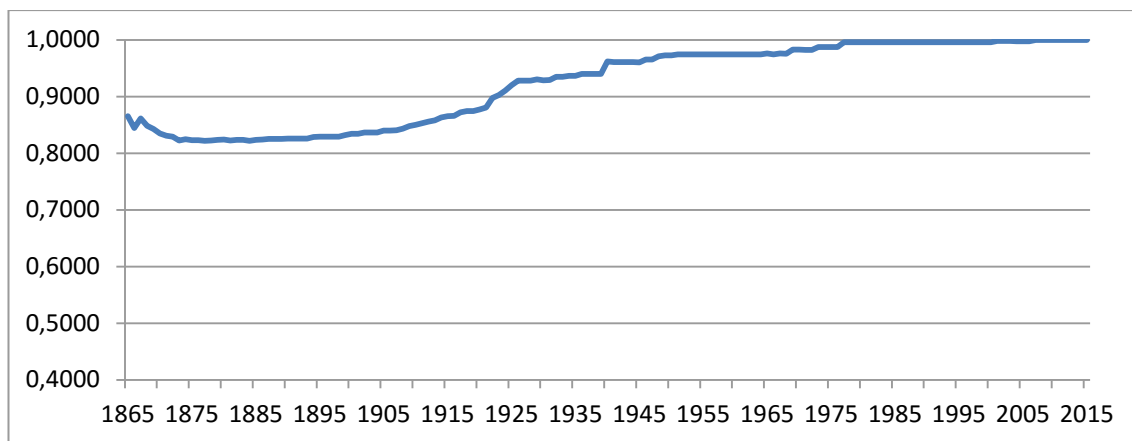
Jokainen jakaantuminen, jossa kunta on lakannut kokonaan olemasta, estää kuntaan merkityn aineiston löytämisen nykyisellä kuntanimistöllä, koska ei voida luoda yhteyttä alkuperäisten ja nykyisten kuntien välille. Sen sijaan, jos jakaantuminen on tapahtunut niin, että kunnan nimi on yhdellä jakaantuneista alueista säilynyt samana, kunnan nimellä pystytään hakemaan aineistoa nykyhetken ja jakaantumishetken väliseltä ajalta.

Tällaisessa tilanteessa myöhemmin tapahtuva yhdistyminen voi kompensoida tapahtuneen jakaantumisen, jolloin vanhempikin aineisto on taas löydettävissä koko ajalta, vaikka yhdistymisessä kunnan nimi muuttuisi. Kaikki tällaiset tapaukset on menetelmän toimivuutta laskettaessa otettu huomioon. Nimenmuutokset pystytään helposti ottamaan

huomioon tietokannan muutostaulun avulla laajentamalla haku koskemaan myös muutunutta nimeä.

Olen laskenut saannin suuruudet kaikista Suomen kunnista oletuksella, että jokaiseen Suomen kuntaan on sijoitettu yksi kohde vuodessa, kuten kappaleessa 7.1 tehtiin saannin perustasoa laskettaessa. Tällä kertaa olen laajentanut haun muutostaulun avulla niin, että nimenmuutokset ja yhdistymiset otetaan haussa huomioon. Tulos, kuten kuvasta 60 nähdään, on suorastaan hämmästyttävän hyvä. Vuoteen 1923 saakka saanti pysyy yli 90 %:ssa ja on koko kuntahistorian ajan yli 82 %, kun ilman muutostaulua saanti laskee jo vuonna 2004 alle 70 %:n ja on huonoimmillaan 52 %.

Todelliselle aineistolle saatavat saantiarvot riippuvat toisaalta aineiston ajallisesta jakaantumisesta ja toisaalta siitä, mistä kunnista aineisto on peräisin. Tässä laskettiin arvot oletuksella, että aineisto jakaantuu tasaisesti kaikkien Suomen kuntien kesken, mutta yhtä hyvin voidaan laskenta tehdä mille aineistolle tahansa, jolle aika- ja sijaintitiedot ovat saatavilla.



Kuva 60. Vuosittain laskettu saanti käyttäen tietokannassa nimenmuutokset ja yhdistymiset sisältävää muutostaulua oletuksella, että aineisto jakaantuu tasaisesti Suomen kuntien kesken.

Tulosta jonkin verran huonontaa todellisesta se, että osa relevanteista kohteista jää laskematta tilanteessa, että jakaantuminen on tapahtunut niin, että kunnan nimi on yhdellä jakaantuneista alueista säilynyt samana. Jakaantumista edeltävää aikaa ei ole laskussa otettu lainkaan huomioon, vaikka osa jakaantuneella alueella sijainneista kohteista voivat olla täysin relevantteja. Tämä puute johtuu siitä, että laskennassa otetaan huomioon vain kokonaisuena relevantit alueet, jolloin vältetään osittaisten alueiden keskinäistä arvottamista esimerkiksi pinta-alojen perusteella.

Tulosta jonkin verran taas todellisuudesta parantavana tekijänä mainittakoon, ettei aluesiirtoja ole otettu laskennassa huomioon. Aluesiirtojen huomioiminen vaatisi arvottamaan sen, millä perusteella siirretyillä alueilla olevien kohteiden painoarvo suhteessa emokuntiin lasketaan: käytetäänkö pinta-alaa, asukaslukua vai jotain muuta tapaa. Pinta-alojen pohjalta laskien siirretyillä alueilla ei ole kovin suurta merkitystä tulokseen, koska siirtyneiden alueiden pinta-alat ovat suhteellisen pieniä.

7.4 Tarkkuus ja saanti SAPO-ontologiaa käytettäessä

Kauppinen ja muut ovat laskeneet julkaisuaan [23] varten SAPO-ontologian käytön vaikutuksen suuruutta tarkkuus- ja saantiarvoihin kahdelle erilaiselle ja eriaikaiselle aineistolle. Laskennassa on oletettu aineiston jakaantuneen tasaisesti annotoinneissa käytettyihin kuntiin, koska ontologiamenetelmän käyttämä todennäköisyyslaskenta perustuu pinta-alasuhteisiin.

Kauppinen ja muiden artikkeli ei ota kantaa siihen, millainen tarkkuus-saanti-tulos aineistolla olisi saatu, jos kuvien todelliset tarkat ottopaikat olisi määritetty ja tehty laskenta sen mukaan, eli aiheuttaako aineiston todellisuudessa enemmän tai vähemmän epätasainen jakaantuminen merkittävää muutosta tulokseen. Todellisten kuvauspaikkojen sijainnin selvittäminen suurille kuvausaineistoille on kuitenkin huomattavan työlästä.

Aineistoina laskennassa käytettiin 12 433 ilmakuvaa vuosilta 1933–1999 ja 3206 Geologian tutkimuskeskuksen luonnontieteellistä valokuvaa vuosilta 1870–1955. Laskennassa käytettiin kirjoitushetken SAPOa vuodelta 2010. Nyt vuonna 2015 saatava tulos voisi olla erilainen, koska laskennassa haun oletetaan tapahtuvan tutkimushetken kuntajaon mukaan ja tällä välillä on tapahtunut runsaasti kuntamuutoksia, pääasiassa yhdistymisiä.

Kuvauspaikaksi aineistolle merkittiin kuvaustietoihin merkitty kuvaushetken mukainen kunta. Ilman laajennettua hakua ilmakuvista löytyisi Kauppinen ja muiden mukaan nykyisen kunnan mukaan haettaessa 67 % ja GTK:n kuvista 59 %. Kun verrataan tulosta Suomen kaikkien kuntien kesken tasaisesti jakaantuvan aineiston perustasoon, joka esiteltiin kappaleessa 7.1, huomataan että näillä aineistoilla saadaan hieman keskimääräistä parempi saanti. Saantihan riippuu tässä tapauksessa paitsi ajasta, myös siitä, miten aineisto jakaantuu eri kuntien kesken. Esimerkiksi, jos kaikki kuvat olisi otettu muuttumattomana pysyneessä kunnassa, vaikka Asikkalassa, tarkkuus ja saanti olisivat kautta historian 100 %.

Kun Kauppinen ja muut laajensivat GTK:n kuvien haun käyttäen pelkästään toisensa täysin peittäviä ajallisia kuntia, saanti parani 71 %:iin tarkkuuden pysyessä edelleen sadassa prosentissa. Jotta saavutettaisiin täysi 100 %:n saanti, otettiin kaikki, myös osittaiset, peittävyys laskennassa huomioon. Tällöin hakutuloksen tarkkuus laski 74 %:n luokkaan. Ilmakuvilla vastaavasti kokonaisilla peittävyyksillä saanti parani 87 %:iin ja kaikki muutokset huomioiden täydellinen saanti saavutettiin tarkkuuden laskiessa 86 %:iin.

Täydellisten peittävyyksien käyttäminen ontologiamenetelmässä vastaa periaatteessa edellisessä kappaleessa kuvattua nimenmuutokset ja yhdistymiset huomioivaa muutostaulumenetelmää. Tulokset eivät kuitenkaan ole täysin vertailukelpoiset, koska:

- Ontologiamenetelmän tuloksissa ei oteta huomioon sitä, että luovutettujen alueiden kohteita haettaisiin luovutushetken eikä nykyisen kunnan nimen mukaan.
- Ontologiamenetelmän tulosten laskenta-ajan jälkeen on tapahtunut runsaasti tuloista mahdollisesti parantavia kuntaliitoksia.
- Ontologiamenetelmän vaikutusta arvioitaessa käytettiin aineistoa, jonka jakaantumista eri vuosille ja eri kuntien kesken ei kuvata tarkemmin. Muutostaulumenetelmän tulos taas on kaikkien Suomen kuntien keskimääräinen arvo.

7.5 Ajallisen paikkatiedon hallintaan liittyvä kysely

Muuttuvan paikkatiedon hallintamenetelmiä tutkiessa on herännyt ajatus selvittää, onko suomalaisissa muisti- ja tutkimusorganisaatioissa koettu paikkatietojen muutosten huomioimista tarpeelliseksi ja jos on, miten ongelmaa on pyritty ratkaisemaan. Kolmas kiinnostava kysymys on, miten asiantuntijoiden mielestä heidän palveluidensa käyttäjät tekevät hakuja. Hakevatko käyttäjät mieluummin nykyisellä paikannimistöllä vai aineiston metatietoihin tallennetulla ”alkuperäisellä” nimistöllä?

Kysely lähetettiin sähköpostitse kymmeneen organisaatioon, joista viidestä saatiin vastaukset. Yhdestä organisaatiosta saatiin kahta eri järjestelmää koskevat vastaukset. Vaikka vastanneiden organisaatioiden lukumäärä on melko pieni, ne ovat kukin aloillaan merkittävimpiä toimijoita Suomessa ja edustavat yhdessä hyvin suurta aineistomäärää. Vastanneet organisaatiot olivat:

- Arkistolaitos
- Geologian tutkimuskeskus (GTK)
- Kotimaisten kielten keskus (Kotus)
- Luonnontieteellinen keskusmuseo, kehitteillä oleva kokoelmatietojärjestelmä (Kotka)
- Luonnontieteellinen keskusmuseo, Eläinmuseon selkäsarkaisten näytteiden kokoelmien tietokanta (SELMA)
- Museovirasto

Useita aineistojen kuvailujärjestelmiä ollaan valtakunnallisesti yhtenäistämässä. Esimerkiksi luonnontieteelliset museot ovat siirtymässä kymmenistä erilaisista tietokannoista yhteiseen kokoelmajärjestelmään (Kotka) ja Arkistolaitos on taas mukana rakentamassa uutta hakemistotietokantaa (AHAA) monien aiemmin käytettyjen järjestelmien korvaajaksi. Näihin liittyvät vastaukset edustavat tätä kautta vastaajaorganisaatioitaankin laajempaa osaa Suomen muistiorganisaatioista ja tietoaaineistoista.

Kyselyssä esitettiin seuraavat kysymykset (*kursiivilla*). Kunkin kysymyksen perässä lyhyt kuvaus, mitä kysymyksellä halutaan saada selville.

1. *Kuvaile, miten tallennatte aineistojen paikkatiedot metatietoihin. Mitä metatietokenttiä on käytössä ja mitä tietoja niihin syötetään?*

Käytettävissä olevat metatietokentät kertovat, millä tarkkuudella organisaatio tallentaa paikannimistötietoja, tallennetaanko koordinaattitietoja ja jos tallennetaan, millä tarkkuudella (piste-, rajauslaatikko- vai polygonitieto). Eniten käytetyt paikkatyypit ovat sellaisia, joihin muutosten hallinta kannattaa sitä tarvittaessa ensimmäisenä kohdistaa.

2. *Mikä on teille tärkein paikkatieto, joka tallennetaan?*

Kysymyksellä halutaan selvittää, mikä paikkatieto vähintään tallennetaan kullekin aineistolle.

3. *Jos teillä on tallennettavana vanhempaa aineistoa, käytättekö sen kuvailussa aineiston valmistusajan kuntia, nykyistä kuntajakoa vai jotain muuta periaatetta, minkä ajan mukaan kuntatieto merkitään?*

Tämä tieto kertoo osaltaan, miten organisaatiossa hallitaan ja tarvittaessa pysyttäisiin hallitsemaan paikkatiedon muutoksia ja miten aineistoa ensisijaisesti haetaan.

4. *Onko teillä käytössä mitään menetelmää hallita paikannimissä tapahtuvia muutoksia, esimerkiksi kuntien jakaantumisia, liitoksia ja nimenmuutoksia?*

Vastaus kysymykseen kertoo suoraan, onko paikannimien muutosten hallintaan yritetty kehittää mitään menetelmää. Jos on, organisaatiossa on menetelmälle myös ilmeinen tarve.

5. *Jos ei ole, olisiko sellainen tarpeen? Jos on, niin onko käyttämänne menetelmä toiminut riittävän hyvin tarpeisiinne?*

Jos menetelmää ei ole käytössä, sille voi kuitenkin olla tarvetta. On myös mahdollista, että käytössä oleva menetelmä on täysin riittävä organisaation tämänhetkisiin tarpeisiin.

6. *Jos haet aineistoa kunnan perusteella, haetko sitä nykyisen kuntajaon perusteella vai käytytkö lakaneiden kuntien nimiä? Miksi niin? Jos haet vanhoilla nimillä, mistä saat tietää, millainen kuntajako on ollut haluamallasi hetkellä?*

Kysymykset on lähetetty organisaationsa paikkatiedon hallinnasta jossakin roolissa vastaaville henkilöille eli organisaationsa asiantuntijoille. Tällä kysymyksellä halutaan tietää, miten asiantuntija itse hakee tai hakisi aineistoa.

7. *Millä kunnannimistöllä olettaisit ihmisten yleensä hakevan aineistoanne, nykyisellä vai vanhalla? Miksi?*

Asiantuntijan näkemys, kuinka organisaation asiakkaat hänen mukaansa hakevat heidän aineistoaan.

8. *Voiko aineistoanne hakea koordinaattitietojen perusteella esimerkiksi karttapohjalta?*

Kysymyksellä selvitetään mahdollisuutta koordinaattien perusteella tehtäviin hakuihin.

9. *Jos voi, millä tavalla vanhempaan aineistoon on saatu koordinaattitiedot? Paikannimien perusteella georeferoimalla?*

Jos koordinaattitiedot on muodostettu paikannimien kautta georeferoimalla, niiden sijaintitiedon oikeellisuus ja tarkkuus riippuu siitä, miten hyvin paikannimistö on kyetty yksilöimään ja minkä tyyppiset koordinaattitiedot on ollut käytävissä (piste-, rajauslaatikko- vai polygonitiedot). Tarkkuuteen vaikuttaa myös se, onko georeferointi tehty alkuperäistä kuvailutietoa alueellisesti vastaavaan ajalliseen paikkaan.

Tallennettavista paikkametatiedoista eniten käytetty oli kunta. Sitä varten ovat käytössä tallennuskentät kaikkien organisaatioiden järjestelmissä, joissa ylipäänsä käyte-

tään paikannimistöä. Kenttien niminä kuntatiedolle käytettiin myös pitäjää (Kotus) ja paikkakuntaa (Arkistolaitos). Kotuksessa pitäjillä tarkoitetaan sotia edeltävää kuntajakoa, jolloin se nimenkin puolesta on helppo erottaa nykyisestä kuntajaosta. Haastateluista organisaatioista Kotuksen aineistotietokannassa on jo käytetty SAPO-ontologiaa rajapinnan kautta silloin kun se oli toiminnassa. Arkistolaitoksen tulevassa AHAA-tietokannassa ja mahdollisesti myös luonnontieteellisten museoiden tulevassa keskitetyssä tallennusjärjestelmässä Kotkassa tullaan käyttämään SAPOa, kun se julkaistaan Finto-palvelun kautta.

Kaikki organisaatiot pyrkivät säilyttämään alkuperäisen kuntatiedon tietojärjestelmissään. Muuttuneita kuntatietoja hallitaan eri organisaatioissa ja organisaatioiden sisällä monilla eri tavoilla:

- Arkistolaitos: ”Tarvittaessa/mahdollisuuksien mukaan ajallisia muutoksia on kuvattu tekstikentissä.” Apuna on käytetty manuaalisia kortistoja, kirjallisuutta sekä internethakua.
- Geologian tutkimuskeskuksessa on käytetty ensisijaisena paikkatietona koordinaatteja. Valokuvatietokannoissa on käytetty tässä työssä kappaleissa 5 ja 6 kuvattuja menetelmiä.
- Kotimaisten kielten keskuksessa on käytetty murteiden jaottelussa ja arkistokoelmien järjestämisessä vanhaa pitäjäjakoja, johon alan asiantuntijat ovat tottuneet. Suomen kielen nauhoitearkistossa on käytetty kuntamuutostietoja uudemmilla kunnilla tehtävien hakujen helpottamiseksi. Kokoelmätietokannassa kehitettiin SAPOn käyttöä muutosten hallinnassa, kunnes SAPOn rajapinta ONKI-palvelussa lakkasi toimimasta. ”Uusissa ja karttuvissa kokoelmissa saatetaan käyttää nykyistä tai ainakin uudempaa kuin sotia edeltänyttä kuntajakoa, ja yhden arkistokokonaisuuden sisälläkin saattaa olla pientä variaatiota.”
- Luomus: ”Vaihtelee kokoelmittain ja organisaatioittain”. ”Osa ihmisistä on luottanut vanhaan kuntajakoon, esim. 70-luvulla Societas pro Fauna et Flora Fennican julkaisemaan karttaan.” Selma-tietokannassa on Suomen kunnat taulukko, johon on koottu tietoa kuntien nimi- ja aluemuutoksista.
- Museovirasto: ”Kulttuuriympäristön suojelukohteille tallennetaan ja päivitetään joka vuosi muuttuneet kuntatiedot. Kokoelmahallinnan osalta päivitystyötä ei ole tehty.”

Tarve hallinnollisten alueiden muutosten huomioimiselle on siis ollut ja on edelleen ilmeinen kaikissa haastatelluissa organisaatioissa. Tämä kävi ilmi myös vastauksissa kysymykseen viisi, jossa asiaa suoraan tiedusteltiin. Tiivistetysti erilaisia tapoja hallita muuttuvia tietoja haastatelluissa organisaatioissa ovat olleet:

1. Haettaessa käytetään apuna kortistoja, kirjallisuutta, internetiä tai omaa tietämystä hallinnollisten alueiden muutoksista.
2. Hallinnollinen aluejako on ”sementoitu” johonkin sopivaan ajanhetkeen, jonka mukaan aineistoa kuvaillaan ja haetaan.
3. Käytetään omaa aineistolle soveltuvaa aluejakoa, joka ei riipu hallinnollisista rajoista (esimerkiksi eliömaakuntia tai havaintoruutuja luonnontieteellisille aineistoille tai karttalehtijakoa kartoitusaineistolle)
4. SAPO-ontologiaa on jo käytetty, vaikka sen ylläpidosta ei vielä ole aiemmin ollut päätöstä.
5. Kulloinkin käytössä oleva kuntajako päivitetään aineistolle vuosittain kuitenkin säilyttäen myös alkuperäiset tiedot.
6. Lisätieto- tai vastaaviin tallennuskenttiin kirjoitetaan tietoa hallinnollisten alueiden muutoksista.

Kohtien kaksi ja kolme menetelmät olivat uusia verrattuna aiemmin tässä työssä kuvattuihin tapoihin hallita muuttuvia paikkatietoja.

Aineistoon tapahtuvia hakuja haastatellut asiantuntijat olettavat tapahtuvan sekä nykyisellä että alkuperäisellä kuntanimistöllä. Peruskäyttäjien arveltiin käyttävän pääasiassa nykyistä kuntajakoa ja asiantuntija-asiakkaiden tarpeen mukaan joko alkuperäistä tai nykyistä jakoa. Arkistolaitoksen haastateltava kuvasi: ”Itsenäisesti hakiessaan asiakkaat käyttänevät hakusanana sitä kunnan nimeä, jonka ovat oppineet tuntemaan. Valistuneemmat aineiston hakijat osaavat ottaa huomioon muutokset kunnan nimissä eri ajankohtina.” Tämä voi tarkoittaa sitäkin, että osa käyttäjistä hakee nimillä, jotka ovat ”äskettäin” muuttuneet, esimerkiksi puoli vuotta sitten toiseen kuntaan yhdistyneen kunnan nimellä, koska ei tiedä tapahtuneesta kuntaliitoksesta tai ota sitä huomioon.

Edelliseen viittaa myös GTK:n haastateltu: ”Ehkä joku vanhemman polven edustaja voisi käyttää aivan viime vuosina vanhaksi mennyttä kuntanimeä.” Tästä syystä nykyisillä kunnilla tapahtuvaan hakuun olisi hyvä tulla tuloksena myös lakanneiden kuntien aineisto, jos niiden nimellä haetaan. Hakujärjestelmä on siksi parempi toteuttaa niin, että haku tapahtuu joko alkuperäisellä kunnalla tai alkuperäisistä nykyisiin kuntiin laajennettuna, kuten kappaleessa 6.3 esittelemässäni kuntomuutostaulumenetelmässä tapahtuu eikä joko nykyisillä tai alkuperäisillä kunnilla.

Arkistolaitosta lukuun ottamatta kaikissa organisaatioissa on kerätty ainakin osalle aineistoja myös koordinaattitietoja. Tuleva AHAA-järjestelmä tulee mahdollistamaan koordinaattien tallennuksen myös Arkistolaitoksen aineistoille. Koordinaatteja vanhemmalle aineistolle saadaan georeferoimalla paperikartoille tehtyjä merkintöjä (Museovirasto ja Kotus) tai paikannimien perusteella hakemalla käsityönä (GTK:n valokuvat ja Luomus). Arkistolaitoksessa ja Luomuksessa on myös harkittu mahdollisuutta automaattiseen georeferointiin paikannimien perusteella. Se edellyttää tietenkin sitä, että on saatavissa paikannimiä vastaavat koordinaattitiedot riittävällä tarkkuudella ja oikealta ajalta.

8 Pohdinta

8.1 Tutkimuksen tuloksista

Muutostyypit

Tässä työssä esiteltiin Suomen kunnissa tapahtuneet 12 erityyppistä muutosta, jotka voivat vaikuttaa tietoaaineistojen löydettävyyteen. Missään aiemmassa tutkimuksessa ei ole kuvattu näin montaa hallinnollisten alueiden muutostyyppiä. Esimerkiksi Gregory ja muut [6] esittelevät Isosta-Britanniasta vain neljä perusmuutostyyppiä. He olivat havainneet kyllä monimutkaisempiakin muutoksia, mutta totesivat että muut muutokset voidaan kuvata neljän perustyyppin avulla. Olemme kuitenkin havainneet, että ainakin Suomessa on hyödyllistä käyttää useampia muutostyyppejä.

Muutostyyppien vähyys Isossa-Britanniassa verrattuna Suomeen johtuu toisaalta siitä, että Gregory ja muut eivät mainitse perustamista ja lakkauttamista muutostyypeiksi ja toisaalta siitä, että Ison-Britannian tutkimusalueelta puuttuvat Suomen alueluovutuksen ja aluesaamisen kaltaiset muutokset. Sama tilanne on muutostyypeillä aluevuokraus ja aluevuokrauksen päättymisen.

Uudet julkaisussamme [56] ensimmäistä kertaa esitellyt muutostyypit aluesiirto-perustaminen ja aluesiirtojakaantuminen ovat useantyyppisten muutosten samanaikaisia yhdistelmiä. Aluesiirto-perustamisessa tapahtuu samanaikaisesti useita aluesiirtoja ja uuden alueen perustaminen siirtyville alueille. Aluesiirtojakaantumisessa puolestaan hallinnollinen alue lakkautetaan ja sen alueet siirretään useaan olemassa olevaan hallinnolliseen alueeseen. Näitä muutostyyppejä ei voida esittää Gregoryn ja muiden esittämien perusmuutostyyppien (aluesiirto, nimenmuutos, jakaantuminen ja yhdistyminen) avulla. Tarvitaan aluesiirtojen lisäksi samanaikainen hallinnollisen alueen lakkauttaminen tai perustaminen. Väitänkin, että käytännöllisin tapa kuvata tietentyypistä samalla hetkellä tapahtuvaa muutosjoukkoa on esittää se omana muutostyyppinä.

Muutostenhallintamenetelmät

Tässä työssä esiteltiin kolme erilaista menetelmää, joiden avulla hallinnollisten alueiden muutoksia pystytään mallintamaan niin, että tietojärjestelmistä tapahtuvien hakujen saanti paranee. Kaikilla esitetyillä menetelmillä on omat käyttötarpeensa.

Ontologiamenetelmä on tässä esitellyistä ainoa, joka ei ole sisältötyötä lukuun ottamatta kirjoittajan itsensä kehittämä. Menetelmän toimivuuden kannalta on hyväksi, jos haettava aineisto on alun perin kuvailtu ontologian avulla. Silloin samannimiset paikat voidaan erottaa toisistaan eivätkä ne tuota virheitä hakutuloksiin. Ontologiamenetelmällä voidaan saada suuri parannus aineistojen löydettävyydelle ilman hakutarkkuuden huonontumista tai tarvittaessa täydellinen saanti ilman että tarkkuus oleellisesti laskee, jolloin hakutulokset voidaan järjestää relevanssin todennäköisyyden mukaiseen järjestykseen. Oletuksena tällöin on, että aineisto on jakaantunut suhteellisen tasaisesti haettavalle alueelle.

Menetelmän toimivuus on testattu Kulttuurisampo-portaalissa, mutta yhtään tutkimusmaailman ulkopuolista menetelmää käyttävää sovellusta ei ole vielä toteutettu. Peittävyystietojen käyttöönotto ja ylläpidon järjestäminen vaatii sovelluksen kehittäjältä jonkin verran osaamista. Suurempi ongelma on se, että todennäköisyyksien laskennassa tarvittavia peittävyystietoja ei tällä hetkellä ole saatavissa. Toivon mukaan tämä ongelma korjaantuu, kun SAPO julkaistaan Finto-palvelussa ja sen ylläpito järjestetään.

Kategoriamenetelmä soveltuu yksittäisen organisaation sisäisen tietojärjestelmän käyttöön. Sen avulla voidaan saavuttaa täysi saanti ja tarkkuus kaikelle aineistolle. Suurimmat rajoitukset menetelmässä ovat se, että haku voi tapahtua pelkästään kategorioita valitsemalla ilman hakusanakenttää ja sen soveltuminen sellaisenaan vain yhteen kaupalliseen aineistohallintaohjelmaan.

Tietokannan muutostaulumenetelmä parantaa aineistojen löydettävyyttä lähes yhtä paljon kuin ontologiamenetelmä. Sen etuna on erittäin yksinkertainen käyttöönotto normaalissa verkkopalvelussa. Tietokantaan tarvitsee lisätä vain yksi taulu, joka sisältää tarvittavat hakulaajennukset muuttuneille kunnille. Käyttöliittymään lisätään valinta, käytetäänkö muutostaulua vai ei. Valinta määrittää sen, haetaanko alkuperäisellä kuntanimellä vai nykyisellä. Ongelmana menetelmässä on se, ettei se erota keskenään samanimisiä paikkoja, toisin kuin ontologia. Toinen puute on se, että osaa jakaantuneista alueista ei pystytä menetelmän avulla löytämään lainkaan nykyisellä nimellä. Toisin sanoen sataprosenttista saantia ei kyetä saavuttamaan.

Mikä menetelmä soveltuu missäkin tilanteessa käytettäväksi? Lyhyesti voidaan suositella:

- Kun käytetään aineistohallintaohjelmaa, jossa on käytettävissä loogisten operaattoreiden avulla määritettäviä kategorioita (esimerkiksi IMatch), voidaan käyttää kategoriamenetelmää.
- Kun aineiston kuvailussa käytetään ontologioita, myös hakujen laajentamiseen kannattaa käyttää ontologiamenetelmää. Samoin on, jos halutaan löytää maksimaalinen osa relevanteista kohteista.
- Jos halutaan lisätä tietojärjestelmään alkuperäisen kuntatiedon lisäksi mahdollisuus hakea nykyisellä kuntanimistöllä, muutostaulumenetelmä on helposti toteutettavissa ja parantaa saantia huomattavasti.

8.2 Katsaus tutkimuskysymyksiin

Palataan kappaleessa 1 esitettyihin tutkimuskysymyksiin ja katsotaan, saatiinko niihin vastauksia:

1. Millaisia aluemuutoksia Suomen hallinnollisissa alueissa on tapahtunut, ja miten tapahtuneet muutokset vaikuttavat tietojärjestelmiin tallennetun tiedon löydettävyyteen?

Suomen hallinnollisissa alueissa on aikavälillä 1865–2015 tapahtunut 12 erityyppistä muutosta, jotka vaikuttavat alueiden nimiin tai rajoihin. Muutostyypit (perustaminen, lakkauttaminen, jakaantuminen, yhdistyminen, aluesiirto, nimenmuutos, aluesiirtoperustaminen, aluesiirtojakaantuminen, alueluovutus, aluesaaminen, aluevuokraus ja aluevuokrauksen loppuminen) esiteltiin kappaleessa 3.1. Lisäksi esiteltiin kaksi muutostyyppiä, jotka koskevat vanhempia aikoja (muodostuminen ja rajanveto).

Muutosten vaikutukset muutoksen kohteena olevien alueiden pinta-aloihin käsiteltiin kappaleessa 3.1 muutosmuotojen esittelyn yhteydessä. Pinta-alasuhteita voidaan käyttää eriaikaisten alueiden pinta-alallisten päällekkäisyyksien laskemisessa ja sitä kautta myös aineistojen sijaintitodennäköisyyksien arvioinnissa. Kappaleessa 3.2 pohdittiin eri muutostyyppien ajallisen sijoittumisen vaikutusta tietojen löydettävyydelle. Lopputuloksena todettiin että mallintamisen kannalta on onnekasta se, että jakaantumiset ovat Suomessa tapahtuneet pääosin ennen yhdistymisiä. Osa jakaantumisista jopa kompensoituu myöhemmän yhdistymisen ansiosta.

2. Mitä erilaisia menetelmiä voidaan käyttää, jotta paikkatiedossa tapahtuvat muutokset tulevat otetuksi huomioon tietoja tietojärjestelmiin tallennettaessa ja niistä haettaessa, ja miten menetelmät vaikuttavat tietojärjestelmiin tehtävien hakujen tarkkuuteen ja saantiin?

Kappaleessa 2 kuvattiin kirjallisuudessa esitettyjä menetelmiä paikkatiedon ja siinä tapahtuneiden muutosten hallintaan. Menetelmät jaettiin nimistöluettelo-, ontologia- ja GIS-menetelmiin. Kappaleessa 4 esiteltiin Suomen ajallinen paikkaontologia, jonka kehittämisessä olen ollut mukana. Kehittämäni kategoriamenetelmä Geologian tutkimuskeskuksen valokuvatietokantaa varten esiteltiin kappaleessa 5 ja tietokannan muutostaulun käyttö muutosten hallintaan kappaleessa 6.

Eri menetelmien vaikutusta tietojärjestelmiin tehtyjen hakujen tarkkuuteen ja saantiin pohdittiin kappaleissa 7.1 – 7.4. Kaikki tutkitut menetelmät paransivat merkittävästi saantia ilman tarkkuuden laskua. Alueiden muutoksia mallintavien menetelmien käytöstä on siis pelkkää etua tilanteessa, että käyttäjät hakevat aineistoja nykyisen kuntajaon perusteella.

3. Miten suomalaisissa muisti- ja tutkimusorganisaatioissa tallennetaan ja haetaan sanallista paikkatietoa ja otetaanko paikkatiedon muuttuminen jotenkin huomioon?

Haastatelluissa organisaatioissa tallennettiin paikkatiedot pääasiassa aineiston syntyajan mukaisella paikannimistöllä. Poikkeuksena oli Kotimaisten kielten keskus, jossa tietyt (esimerkiksi murteiden sanakokoelman) aineistot on luokiteltu tietynhetkisen vanhan aluejaon mukaan. Myös Luonnontieteellisessä keskusmuseossa on aineistoja, jotka on alun perin kuvailtu ”lukitun” kuntajaon mukaisesti. Osalla aineistoa ei ole organisaatioissa säilynyt tietoa siitä, minkä kuntajaon mukaan ne on kuvailtu. Hakua varten organisaatioissa oli kehitetty monentyyppisiä menetelmiä ja apuvälineitä, jotta aineisto löytyisi muuttuneista paikkatiedoista huolimatta. Kaikissa organisaatioissa koettiin hyödylliseksi se, että aineistoja voitaisiin hakea sekä alkuperäisen että muuttuneen paikannimistön perusteella tarpeen mukaan. Kyselytutkimuksen yksityiskohtaisemmat tulokset esiteltiin kappaleessa 7.5.

8.3 Tulevaisuudennäkymiä

Ajallisen paikkatiedon tutkimus on ollut viime vuosina kiihkeässä kehitysvaiheessa. Myös erilaiset paikkatiedon visualisointitavat ovat kokeneet parannuksia ja aiemmin tutkijoiden käytössä olevia menetelmiä ollaan tuomassa suuren yleisön käyttöön. Tekevässäni haastattelussa kävi ilmi, että monissa suurissa kokoelma- ja arkistoaineistoja omaavissa organisaatioissa ollaan uusimassa kokoelmanhallintajärjestelmiä lähiaikoina. Samassa yhteydessä olisi luontevaa ottaa käyttöön tutkimustyönä kehitettyjä menetelmiä, joilla parannetaan merkittävästi aineistojen löydettävyyttä. Merkittävä askel Suomessa tähän suuntaan on SAPO-ontologian julkaisemisen ja ylläpidon toteutuminen Kansalliskirjaston Finto-palvelun kautta.

Kansalliskirjasto on julkaissut kesällä 2015 kirjoittamamme raportin Suomen ajallisen paikkaontologian kehittämis ehdotuksista [56]. Siinä kuvataan yksityiskohtaisesti, miten SAPOa kannattaisi mielestämme kehittää. Suurimpina toiveina käyttäjien taholta on SAPOn ylläpidon vakiinnuttamisen lisäksi ontologian ajallisen ulottuvuuden lisää-

minen kauemmas historiaan ja ontologian kattavuuden laajentaminen muihin paikka-tyyppeihin, esimerkiksi kyliin ja kaupunginosiin.

Tätä työtä varten tekemästani haastattelusta kävi ilmi myös suuri tarve historiallisten paikkojen sijainti- ja rajatietojen näyttämiseen ja käyttämiseen karttakäyttöliittymissä. Tähän tuo osaltaan helpotusta Semanttisen laskennan tutkimusryhmän vuonna 2014 alkanut ja edelleen käynnissä oleva hanke *Historiallisten paikkojen kansallinen nimistöpalvelu*¹⁶ (HIPLA). Valmistuessaan palvelu kerää yhteisöllisesti eriaikaista paikannimistöä ja sijaintitietoja [11]. Suomen ajallista paikkaontologiaa käytetään osana nimistöpalvelua ja HIPLAn avulla siihen voidaan kytkeä kaivattuja muita paikkahierarkian tasoja. HIPLAssa karttakäyttöliittymällä on merkittävä rooli sekä paikkojen lisäämisessä että hakemisessa.

HIPLAn kaltaiset ontologioita hyödyntävät ja GIS-teknologioita käyttävät palvelut tulevat yleistymään. Muistiorganisaatioiden yhteistyössä kehitettävät järjestelmät tulevat korvaamaan yksittäisten organisaatioiden omat järjestelmät. Eri tutkimusprojekteissa kehitettyjä menetelmiä otetaan käyttöön ja tarpeen mukaan kehitetään edelleen käytännössä ilmenevien tarpeiden mukaisesti.

¹⁶ <http://hipla.fi>

9 Yhteenveto

Paikannimistö on monissa, varsinkin historiallista tietoa sisältävissä tietojärjestelmissä, eniten käytetty tapa määrittää paikkoja. Hallinnollisen aluejaon mukaan on koottu erilaisia tilastotietoja ja monia aineistoja on kuvailtu käyttäen paikkatietona esimerkiksi lääni- ja kuntanimiä. Keräämäni aineiston mukaan Suomen kunnissa on tapahtunut yhteensä 1101 muutostapahtumaa, jotka koskevat 1736 kuntaa. Muutokset olen luokitellut kahteentoista erilaiseen tyyppiin: perustaminen, lakkauttaminen, jakaantuminen, yhdistyminen, aluesiirto, nimenmuutos, aluesiirtooperustaminen, aluesiirtojakaantuminen, alueluovutus, aluesaaminen, aluevuokraus ja aluevuokrauksen loppuminen.

Tekemäni haastattelututkimuksen mukaan aineistoja haetaan sekä alkuperäisellä että nykyisellä paikannimistöllä. Nykyisellä nimistöllä haettaessa paikannimissä ja alueiden rajoissa tapahtuneet muutokset huonontavat merkittävästi aineistojen löydettävyyttä. Tällöin ennen vuotta 2004 ajoittuvasta aineistosta löytyy alle 70 % ja ennen vuotta 1973 aineistosta enää alle 60 %. Haastattelussa ilmenikin organisaatioiden tarve muuttuneilla alueilla sijaitsevan aineiston löydettävyyden parantamiseen nykyisillä kunnilla haettaessa.

Löydettävyyttä voidaan parantaa käyttämällä tässä työssä esiteltyjä menetelmiä. Kaikkein kattavimmin erityyppisiä muutoksia voidaan mallintaa käyttämällä ajallista paikkaontologiaa. Suomen ajallisessa paikkaontologiassa (SAPO) kaikki löydetty muutos-tyypit on mallinnettu. Tässä työssä kuvataan Kauppinen ja Hyvösen [22] esittelemä menetelmä, jonka avulla SAPO-ontologian mukaan annotoidusta aineistosta voidaan laskea eriaikaisten ajallisten kuntien peittävyys. Peittävyysien avulla voidaan laskea jonkin ajallisen kunnan alueelle tasaisesti jakaantuneen aineiston todennäköisyydet sijaita jonkin toisen ajallisen kunnan alueella. Lukua voidaan käyttää esimerkiksi hakutulosten järjestämisessä osuvuuden mukaan hakutulostilastassa.

Geologian tutkimuskeskuksen valokuvatietokannassa käytetään itse kehittämäni kategoriamenetelmää, jossa kaikki erityyppiset muutokset on mallinnettu käyttäen loogisia operaattoreita AND ja OR. Kuvat sijoitetaan alkuperäistä kuvausohjetta vastaavaan ajalliseen kategoriaan ja loogisia operaattoreita sisältävät kategoriat kopioivat kuvan mahdollisten välivaiheiden kautta nykytilannetta vastaavaan kuntakategoriaan. Näin kuvaa voidaan hakea alkuperäisestä tai nykyisestä kuntakategoriasta. Tarvittaessa kuva löytyy myös eriaikaisista välivaiheiden kategorioista.

Olen kehittänyt myös Suomen kuntien nimenmuutokset ja yhdistymiset mallintavan menetelmän, jota voidaan käyttää relaatiotietokannoissa. Tietokantaan lisätään muutokset sisältävä taulu, joka laajentaa nykyisillä kuntanimillä tapahtuvat haut kattamaan myös vanhemmat samalla alueella sijainneet kunnat. Menetelmää on käytetty Geologian tutkimuskeskuksen Hakku-hakupalvelun valokuvaosiossa, jossa käyttäjillä on mahdollisuus hakea aineistoja valinnan mukaan alkuperäisillä kuntanimillä tai laajennettuna myös nykyisillä kuntanimillä.

Kaikki esitelty kolme menetelmää parantavat vanhempaan aineistoon kohdistuvien hakujen saantia merkittävästi. Kategoriamenetelmässä saanti on 100 %, mikäli koko aineisto on pystytty sijoittamaan oikeisiin lähtökategorioihin, joka voi vaatia kuntatietoa tarkempaa paikkatietoa. Pelkästään nimenmuutokset ja yhdistymiset mallintavalla muutostaulumenetelmällä saavutetaan yli 82 %:n saanti kautta koko kuntahistorian, jos aineisto jakaantuu keskimäärin kaikkien Suomen kuntien kesken. Sama tulos saavutetaan myös ontologiamenetelmällä vastaavassa tilanteessa. Jos kaikki tapahtuneet muutokset otetaan huomioon, tulos paranisi mahdollisesti vielä edellisestä. Kaikki muutokset huomioimalla on ontologiamenetelmällä mahdollista saavuttaa tarvittaessa 100 % saanti, mutta tarkkuus putoaa siinä tapauksessa jonkin verran.

Lähdeluettelo

- [1] Bernhardsen, T. (2002). *Geographic information systems: an introduction* (3. painos). New York: John Wiley & Sons.
- [2] Brickley, D., & Guha, R. V. (2014). RDF Schema, W3C Recommendation 25 February 2014. Saatavissa: <http://www.w3.org/TR/rdf-schema/>
- [3] Carroll, J. J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., & Wilkinson, K. (2004). Jena: Implementing the Semantic Web Recommendations. *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters - WWW Alt. '04*, 74–83. <http://doi.org/10.1145/1013367.1013381>
- [4] Gantner, F., Waldvogel, B., Meile, R., & Laube, P. (2013). The basic formal ontology as a reference framework for modeling the evolution of administrative units. *Transactions in GIS*, 17(2), 206–226. <http://doi.org/10.1111/j.1467-9671.2012.01356.x>
- [5] Gey, F., Larson, R., Sanderson, M., Joho, H., Clough, P., & Petras, V. (2006). GeoCLEF: the CLEF 2005 cross-language geographic information retrieval track overview, (August). <http://doi.org/10.1007/11878773>
- [6] Gregory, I. N., Bennett, C., Gilham, V. L., & Southall, H. R. (2002). The Great Britain Historical GIS: From maps to changing human geography., 39(1), 37–49. <http://doi.org/10.1179/000870402787288045>
- [7] Grenon, P., & Smith, B. (2004). SNAP and SPAN: Towards dynamic spatial ontology. *Spatial Cognition and Computation*, 1 (March), 69–103. http://doi.org/10.1207/s15427633scc0401_5
- [8] Gruber, T. R. (1993). Technical Report KSL 92–71 Revised April 1993 A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications. *Knowledge Creation Diffusion Utilization*, 5 (April), 199–220. <http://dx.doi.org/10.1006/knac.1993.1008>
- [9] Hill, L. L. (2006). *Georeferencing: The Geographic Associations of Information*. MIT Press.
- [10] Hyvönen, E., Tuominen, J., Kauppinen, T., & Väättäinen, J. (2011). Representing and utilizing changing historical places as an ontology time series. *Geospatial Semantics and the Semantic Web*, 1–25. <http://doi.org/10.1007/978-1-4419-9446-2>
- [11] Ikkala, E., Tuominen, J., & Hyvönen, E. (2015). Contextualizing Historical Places in a Gazetteer by Using Historical Maps and Linked Data. *Submitted*.
- [12] Iptc.org. (2015). Information Interchange Model (IIM). Viitattu 30.10.2015. <https://iptc.org/standards/iim/>

- [13] ISO 16684-1:2012 Graphic technology – Extensible metadata platform (XMP) specification – Part 1: Data model, serialization and core properties (2012). Geneva: International Organization for Standardization.
- [14] ISO 19112:2003 Geographic information - Spatial referencing by geographic identifiers. (2003). Geneva: International Organization for Standardization.
- [15] Janowicz, K. (2006). Towards a Similarity-Based Identity Assumption Service for Historical Places. In *Geographic, Information Science* (Vol. 4197, 199–216). http://doi.org/10.1007/11863939_14
- [16] Janowicz, K., & Keßler, C. (2008). The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science*, 22(10), 1129–1157. <http://doi.org/10.1080/13658810701851461>
- [17] Jones, C. B., Abdelmoty, A. I., & Fu, G. (2003). Maintaining Ontologies for Geographical Information Retrieval on the Web. *Geographical*, 2888, 934–951. Saatavissa: http://www.geo-spirit.org/publications/SPIRIT_maintaining_ontologies.pdf
- [18] Jones, C. B., & Purves, R. S. (2008). Geographical information retrieval, (Elokuu 2015). <http://doi.org/10.1080/13658810701626343>
- [19] Kakkuri, J. (1987). Character of the Fennoscandian land uplift the 20th century. *Special Paper*, 2 (Fennoscandian land uplift). Saatavissa: http://tupa.gtk.fi/julkaisu/specialpaper/sp_002.pdf
- [20] Kauppinen, T. (2010). *GEOSPATIO-TEMPORAL SEMANTIC WEB Doctoral Dissertation. Science And Technology*. Aalto-yliopisto. Saatavissa: <http://lib.tkk.fi/Diss/2010/isbn9789526030593/isbn9789526030593.pdf>
- [21] Kauppinen, T., Henriksson, R., Sinkkilä, R., Lindroos, R., Väättäin, J., & Hyvönen, E. (2008). Ontology-based disambiguation of spatiotemporal locations. *CEUR Workshop Proceedings*, 422. Saatavissa: <http://ceur-ws.org/Vol-422/irsw2008-submission-8.pdf>
- [22] Kauppinen, T., & Hyvönen, E. (2007). Modeling and reasoning about changes in ontology time series. *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*, (chapter 11), 319–338. Saatavissa: <http://kauppinen.net/tomi/kauppinen-et-al-modeling-and-reasoning-2007.pdf>
- [23] Kauppinen, T., Paakkari, P., Mäkelä, E., Kuittinen, H., Väättäin, J., & Hyvönen, E. (2011). Geospatio-Temporal Semantic Web for Cultural Heritage. *Digital Culture and E-Tourism: Technologies, Applications and Management Approaches*, 2010, 48–64. <http://doi.org/10.4018/978-1-61520-867-8.ch004>
- [24] Kauppinen, T., Puputti, K., Paakkari, P., Kuittinen, H., Hyvönen, E., & Väättäin, J. (2009). Learning and Visualizing Cultural Heritage Connections between Places on the Semantic Web. In *Proceedings of the Workshop on Inductive Reasoning and Machine Learning on the Semantic Web (IRMLeS2009)* (pp. 1–12). Heraklion, Crete. Saatavissa: <http://ceur-ws.org/Vol-474/paper7.pdf>

- [25] Kauppinen, T., Väättäinen, J., & Hyvönen, E. (2008). Creating and Using Geospatial Ontology Time Series in a Semantic Cultural Heritage Portal. *S. Bechhofer et Al. (editors): Proceedings of the 5th European Semantic Web Conference 2008 (ESWC 2008), 5021 LNCS*, 110–123. http://doi.org/10.1007/978-3-540-68234-9_11
- [26] Kay, Michael, S. (2007). XSL Transformations (XSLT) Version 2.0, W3C Recommendation 23 January 2007. W3C. Saatavissa: <http://www.w3.org/TR/xslt20/>
- [27] Keßler, C., Janowicz, K., & Bishr, M. (2009). An agenda for the next generation gazetteer: Geographic information contribution and retrieval. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS09*, (c), 91–100. <http://doi.org/10.1145/1653771.1653787>
- [28] Kuotola, Y. (2009). *Kunnan raja*. Tampereen yliopisto. Saatavissa: <https://tampub.uta.fi/bitstream/handle/10024/66526/978-951-44-7841-3.pdf?sequence=1>
- [29] Larson, R. R. (1996). Geographic Information Retrieval and Spatial Browsing. *Geographic Information Systems and Libraries: Patrons, Maps and Spatial Information*, (DI), 81–124. <http://doi.org/10.1145/1008992.1009143>
- [30] Maanmittauslaitos. (2015). Kansalaisen karttapaikka. Viitattu 14.10.2015. <http://kansalaisen.karttapaikka.fi/kartanhaku/paikannimihaku.html?query=hae&hakutapa=paikannimihaku&nimi=Pyh%C3%A4j%C3%A4rvi&kunta=&e=406643&n=7195132&scale=8000000&tool=siirra&styles=normal&lang=fi>
- [31] Manguinhas, H., Martins, B., Borbinha, J., & Siabato, W. (2009). The DIGMAP Geo-Temporal Web Gazetteer Service. *E-Perimtron*, 4(1), 9–24.
- [32] Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. Saatavissa: <http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf>
- [33] McIlraith, S. a., & Son, T. C. (2001). Semantic Web services. *IEEE Intelligent Systems*, 16, 46–53. <http://doi.org/10.1109/5254.920599>
- [34] Palomäki, M. (1987). Human response to the effects of land uplift. *Special Paper*, 2 (Fennoscandian land uplift), 47–50. Saatavissa: http://tupa.gtk.fi/julkaisu/specialpaper/sp_002.pdf
- [35] Peuquet, D. J., & Duan, N. (1995). An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems*, 9(1), 7–24. <http://doi.org/10.1080/02693799508902022>
- [36] Sanastokeskus, T. S. K. (2014). Geoinformatiikan sanasto. Sanastokeskus TSK ry, Helsinki. Saatavissa: <http://www.tsk.fi/tiedostot/pdf/GeoinformatiikanSanasto>
- [37] Sanastokeskus, T. S. K. (2006). Terminologian sanasto. Sanastokeskus TSK ry, Helsinki. Saatavissa: <http://www.tsk.fi/tiedostot/pdf/TerminologianSanasto.pdf>

- [38] Soikkanen, H. (1966). *Kunnallinen itsehallinto - kansanvallan perusta: maalaiskuntien itsehallinnon historia*. Maalaiskuntien liitto.
- [39] Southall, H. (2012). Rebuilding the Great Britain Historical GIS, Part 1. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 45(3), 149–160. <http://doi.org/10.1080/01615440.2011.589774>
- [40] Southall, H. (2012). Rebuilding the Great Britain Historical GIS, Part 2: A Geo-Spatial Ontology of Administrative Units. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 45(3), 119–134. <http://doi.org/10.1080/01615440.2012.664101>
- [41] Southall, H. (2014). Rebuilding the Great Britain historical GIS, Part 3: Integrating qualitative content for a sense of place. *Historical Methods*, 47(1), 31–44. <http://doi.org/10.1080/01615440.2013.847774>
- [42] Southall, H., Mostern, R., & Berman, M. L. (2011). On historical gazetteers. *International Journal of Humanities and Arts Computing*, 5(2), 127–145. <http://doi.org/10.3366/ijhac.2011.0028>
- [43] Southall, H., Von Lünen, A., & Aucott, P. (2009). On the organisation of geographical knowledge: Data models for gazetteers and historical GIS. In *e-science 2009 - Proceedings of the 2009 5th IEEE International Conference on e-Science Workshops* (162–166). <http://doi.org/10.1109/ESCIW.2009.5407970>
- [44] Stuckenschmidt, H. (2009). Spatial Reasoning for the Semantic Web - Use Cases and Technological Challenges. *Generalization of Spatial Information*, 1–7. Saatavissa <http://drops.dagstuhl.de/opus/volltexte/2009/2138/>
- [45] Stuckenschmidt, H., & van Harmelen, F. (2005). Information Sharing on the Semantic Web, 252. <http://doi.org/10.1007/b138282>
- [46] Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1-2), 161–197. [http://doi.org/10.1016/S0169-023X\(97\)00056-6](http://doi.org/10.1016/S0169-023X(97)00056-6)
- [47] Tuan, Y.-F. (1975). Place: An Experiential Perspective. *Geographical Review*, 65(2), 151–165. <http://doi.org/10.1007/s13398-014-0173-7.2>
- [48] Valanto, S. (2013). *Musketti-sovelluksen luettelointiohje*. Museovirasto. Saatavissa: <http://www.nba.fi/fi/File/1796/musketin-luettelointiohje-2013-01-04.pdf>
- [49] W3C. (2009). SKOS Simple Knowledge Organization System Primer. Viitattu 3. lokakuuta, 2015. Saatavissa: <http://www.w3.org/TR/skos-primer/>
- [50] Wikipedian Paris-sanan täsmennyssivu (englannin kielellä). Verkkodokumentti. Viitattu 17.8.2015. [https://en.wikipedia.org/w/index.php?title=Paris_\(disambiguation\)&oldid=666088664](https://en.wikipedia.org/w/index.php?title=Paris_(disambiguation)&oldid=666088664)

- [51] Winnige, N. (2000). Germany, Historical Atlas of Prussia. Viitattu 28.9.2015. Saatavissa: http://www.geog.port.ac.uk/hist-bound/project_rep/proj_prussia.htm
- [52] Volz, R., Kleb, J., & Mueller, W. (2007). Towards ontology-based disambiguation of geographical identifiers. *Elements*, 170, 7. Saatavissa: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.9311&rep=rep1&type=pdf>
- [53] Worboys, M., & Duckham, M. (2004). Spatial reasoning and uncertainty. In *GIS: A Computing Perspective* (2nd ed., pp. 323–353). Boca Raton, Florida: CRC Press.
- [54] Väätäinen, J. (2007). *Ajallisesti muuttuvan paikkatiedon hallinta*. Evtek-ammattikorkeakoulu. Saatavissa: <http://seco.cs.aalto.fi/publications/2007/vaatainen-ajallinen-paikkatieto-instyo-2007.pdf>
- [55] Väätäinen, J. (2015). Taulukko, johon on koottu Suomen hallinnollisissa alueissa tapahtuneet muutokset. Espoo.
- [56] Väätäinen, J., & Kauppinen, T. (2015). *SAPOn nykytila, ylläpito ja laajennusmahdollisuudet 2015*. Helsinki: Kansalliskirjasto. Saatavissa: <http://urn.fi/URN:ISBN:978-951-51-1422-8>

Liite 1. Kuntamuutostietojen lähteitä

Kuntahistoriikit

Ahonen, K., Tuunanen, E. ja Elsinen, P. (1986). *Joensuun kaupungin historia II–IV*. Joensuun kaupunki.

Allardt, A. (1925). *Borgå sockens historia till freden i Nystad*. Söderström.

Allardt, A. (1935). *Liljendal sockens historia*. Söderström.

Enbuske, M., Runtti, S. ja Manninen, T. (1997). *Rovaniemen historia vuoteen 1990*. Jyväskylä.

Favorin, M. (1986), *Siuntion historia*. Siuntion kunta.

Kallio, V. (2005). *Karjalohja: Itsenäisen väen pitäjä*. Karjalohjan historyyhdistys.

Kokkonen, H. (1931). *Elimäen pitäjän historia II*. Helsinki.

Kähönen, E. (1985). *Vanha Äyräpää II*. Vanha Äyräpään seura.

Paappa, E., Ropponen, J. (1992) *Jääsken kihlakunnan historia III*. Jääsken kihlakunnan historiatoimikunta.

Manninen, A. (1962). *Kangasniemen historia 2*. Kangasniemen kunta.

Manninen, T. (1995). *Oulun kaupungin historia*. Oulun kaupunki.

Niemelä J. (1989). *Punkalaitumen historia 2*. Punkalaitumen historiatoimikunta.

Rissanen, H. (2007). *Rajan kahta puolen Kangaslahtea*. Kangaslahden nuorisoseura.

Rosenberg, A. ja Selin, R. (1995). *Suur-Tuusulan historia III*. Tuusulan kunta, Keravan kaupunki ja Järvenpään kaupunki.

Rosenberg, A., Lomu, J. ja Manninen, K. (1993). *Mäntsälän historia II*. Mäntsälän kunta.

Sammatti, Elias Lönnrotin kotipitäjä (2002). Kallio V. toim. Sammatin historian tuki. Otava, Helsinki.

Sormunen, I. (1974). *Nurmijärven pitäjän historia III*. Vammala.

Sundström, L., Selén, J., Cederlöf, H. ja Fagerström, C-J. (1988). *Inkoo 650 vuotta*. Inkoon kunta.

Warén, A. (1890). *Keuruun pitäjän historia*. Suomalaisen kirjallisuuden seura, Helsinki.

Vilksa, M. (1926). *Parikkalan historia*. Käkisalmi.

Maakuntahistoriikit:

Wirilander, H. (2008). *Savon historia 4. Uudistuksiin heräävä Savo 1870–1918*. Savon säätiö.

Koivusalo, E. (1987). *Etelä-Pohjanmaan historia VI*. Etelä-Pohjanmaan maakuntaliitto.

Ahlström, C., Haapio, M., Mäkinen, Y. ja Lehmusvaara, I. (1982). *Karjala 3*. Karisto, Hämeenlinna.

Muut julkaisut ja tiedotteet:

Ekko P.O. (1934). Uudenkirkon T.l. kirkkomuistoja.

Iltanen, J. (2004). *Urbes Finlandiae*. Genimap

Paikkala, S., Mikkonen, P., Pitkänen, R. L., Slotte, P. (2007). *Suomalainen paikannimikirja*, Karttakeskus.

Soikkanen, H. (1966). *Kunnallinen itsehallinto - kansanvallan perusta: maalaiskuntien itsehallinnon historia*. Maalaiskuntien liitto.

Suomenmaa 1-9, Maantieteellis-taloudellinen ja historiallinen tietokirja 1906–1931. Tietosanakirja-osakeyhtiö.

Suomenmaa 1-7, Maantieteellis-yhteiskunnallinen tietokirja 1967–1978.

Suomen kartastot 1899, 1910, 1925, 1960 ja 1997

Valtionvarainministeriön tiedotteet kuntamuutoksista

Verkkosivustot:

Verkkosivustojen katsomisaikoja ei ole valitettavasti kirjattu ylös kaikkiin lähteisiin.

Historiasivustot:

<http://www.kajaaninyliopistokeskus.oulu.fi/kainuu/070.htm>

<http://www.luovutettukarjala.fi/>

<http://www.luovutetunetelakarjalanpitajat.fi/> (Katsottu: 15.5.2008)

<http://oulu.ouka.fi/aikamatkaoulu/1800/autonomia.htm>

<http://www.uta.fi/koskivoimaa/kaupunki/1870-00> (Katsottu: 11.4.2004)

Kuntaliitto:

<http://www.kunnat.net/> (Katsottu: 7.1.2009 ja 25.11.2009)

Kuntien omat verkkosivut:

<http://www.hammarland.ax/>

<http://www.kouvola.fi/> (Katsottu: 9.2.2011)

<http://www.taivassalo.fi/>

<http://www.tyrnava.fi/> (Katsottu 4.4.2008)

<http://www.vammala.fi/> (Katsottu: 14.8.2007)

<http://www.vilppula.fi/>

<http://www.varpaisjarvi.fi/> (katsottu: 10.12.2008)

Maanmittauslaitos:

<http://www.maanmittauslaitos.fi/>

Wikipedian artikkelit kuntamuutoksista ja yksittäisistä kunnista:

<https://fi.wikipedia.org/>